

**Reanalysis and Actualisation:
An Investigation of Semantic Factors in the Extension of Nominative Case
to Experiencer Arguments of ME *Liken* and Other Early-English
Impersonal Verbs**

Dissertation zur Erlangung des akademischen Grades
eines Doctor philosophiae (Dr. phil.)

vorgelegt dem Rat der Philosophischen Fakultät
der Friedrich-Schiller-Universität Jena

von

Stephen Zill, M.A.,
geboren am 15. 10. 1982 in Pirmasens

Gutachter:

1. Prof. Dr. Volker Gast (Friedrich-Schiller-Universität Jena)
2. Prof. Dr. Martin. J. Kümmel (Friedrich-Schiller-Universität Jena)
3. Prof. Dr. Thomas Honegger (Friedrich-Schiller-Universität Jena)

Tag der mündlichen Prüfung: 15. 10. 2021

Table of Contents

List of Tables	vii
List of Figures	viii
Acknowledgements	ix
1. Introduction	1
1.1. Impersonal Constructions	1
1.2. Survey of Earlier Literature	3
1.3. Motivation and Outline of the Present Study	8
2. Theoretical Background	11
2.1. Impersonal Verbs	11
2.1.1. Definition	11
2.1.1.1. Subjectless Constructions	12
2.1.1.1.1. Expletive <i>Hit</i>	13
2.1.1.1.2. Clausal Complements	17
2.1.1.1.3. Nominative Complements	19
2.1.1.1.4. Oblique Experiencers	20
2.1.1.1.4.1. Position	20
2.1.1.1.4.2. Coreferential Deletion	21
2.1.1.1.4.3. Raising	23
2.1.1.1.4.4. Verb Agreement	24
2.1.1.1.4.5. Case Marking	26
2.1.1.2. Oblique-Experiencer Verbs	27
2.1.2. Semantic Classification	29
2.1.2.1. Wahlén 1925	29
2.1.2.2. Elmer 1981	32
2.1.2.3. Möhlig-Falke 2012	33
2.1.2.4. Polysemy and Semantic Changes	34
2.1.3. Syntactic Classification	35
2.1.3.1. Ogura 1986	35

2.1.3.2. Elmer 1981	35
2.1.3.2.1. Type S	35
2.1.3.2.2. Type N	38
2.1.4. Theoretical Accounts	40
2.1.4.1. Government and Binding	40
2.1.4.1.1. Single-Entry Approach	40
2.1.4.1.2. Additional Surface Variation	41
2.1.4.1.2.1. Accusative Experiencers	41
2.1.4.1.2.2. Non-Obligatory Arguments	43
2.1.4.1.2.3. PP Constituents	45
2.1.4.1.3. Optionality of Lexical Case	46
2.1.4.2. Case Grammar	47
2.1.4.3. Problems of a Universal Account	48
2.1.4.3.1. Restricted Occurrence of <i>Lician</i>	48
2.1.4.3.2. Other Verbs	50
2.1.4.4. Functional Aspects	53
2.2. Diachronic Change	56
2.2.1. Earlier Accounts	61
2.2.1.1. Reanalysis Theory	61
2.2.1.1.1. Outline	61
2.2.1.1.2. Criticism	64
2.2.1.1.2.1. Empirical Basis	64
2.2.1.1.2.2. Ambiguity and Position of the Experiencer	65
2.2.1.1.2.3. Instantaneity of Change	69
2.2.1.1.2.4. Productivity	70
2.2.1.1.2.5. Lexical Diversity	72
2.2.1.1.3. Summary	74
2.2.1.2. Lexical Functional Grammar	75
2.2.1.2.1. Outline	75
2.2.1.2.2. Criticism	84
2.2.1.3. Principles and Parameters	86
2.2.2. Semantic Hypothesis	90
2.2.2.1. Underlying Theoretical Concepts	91

2.2.2.1.1. Actualisation	91
2.2.2.1.2. Markedness	96
2.2.2.1.3. Semantic Roles	99
2.2.2.1.4. Agentivity	104
2.2.2.1.4.1. Volition	105
2.2.2.1.4.2. Sentience/Perception	110
2.2.2.1.4.3. Causation	111
2.2.2.1.4.4. Summary	113
2.2.2.2. Hypothesis	115
2.2.2.3. Operationalisation	118
3. Methodology	122
3.1. Corpus Selection	122
3.2. Data Extraction	131
3.3. Subsetting	140
3.3.1. Non-finite Forms	140
3.3.1.1. Infinitive	140
3.3.1.2. Present Participle	142
3.3.1.3. Gerund	145
3.3.1.4. Past Participle	146
3.3.2. Passive Constructions	147
3.3.3. Missing Arguments and Null Constituents	149
3.3.3.1. Experiencer	149
3.3.3.2. Stimulus	151
3.3.4. Ambiguous Observations	156
3.4. Annotation of Variables	157
3.4.1. Construction Type	157
3.4.1.1. Case Marking	158
3.4.1.1.1. Pronouns, Nouns, Adjectives	158
3.4.1.1.2. <i>Ye/You</i> Contrast	159
3.4.1.2. Other Criteria	163
3.4.1.2.1. Expletive <i>It</i>	165
3.4.1.2.2. Formulaic Expressions	168

3.4.1.2.3. Case-Identity Constraint	171
3.4.1.2.4. Verb Agreement	173
3.4.2. Independent Variables	175
3.4.2.1. Animacy	175
3.4.2.1.1. Experiencer	176
3.4.2.1.2. Stimulus	180
3.4.2.2. Abstractness	182
3.4.2.2.1. Experiencer	183
3.4.2.2.2. Stimulus	185
3.4.2.3. Category	188
3.4.2.3.1. Experiencer	191
3.4.2.3.2. Stimulus	195
3.4.3. Diachronicity	204
3.4.4. Additional Properties	208
3.4.4.1. Person	208
3.4.4.2. Number	211
3.4.4.3. Gender	212
3.4.4.4. Verbal Features	214
3.4.4.4.1. Grammatical Properties	214
3.4.4.4.2. Modal Auxiliaries	215
4. Results	217
4.1. Liken	219
4.1.1. Database	219
4.1.2. Variables	225
4.1.2.1. Construction Type	225
4.1.2.2. Predictor Variables	229
4.1.2.2.1. Animacy	229
4.1.2.2.2. Abstractness	230
4.1.2.2.3. Category	232
4.1.3. Logistic Regression Model	235
4.1.3.1. Configuration	236
4.1.3.1.1. Periods	236

4.1.3.1.2. Conceptual Variables	238
4.1.3.1.2.1. Levels of Category	239
4.1.3.1.2.2. Levels of Abstractness	242
4.1.3.1.2.3. Selection of Variables	242
4.1.3.1.3. Interactions	245
4.1.3.1.4. Random Effects	252
4.1.3.2. Interpretation	253
4.1.3.2.1. Basic Model	254
4.1.3.2.2. Modifications	256
4.1.3.3. Discussion	258
4.2. Other Verbs	268
4.2.1. Quemen	270
4.2.1.1. Database	270
4.2.1.2. Variables	272
4.2.1.2.1. Construction Type	272
4.2.1.2.2. Semantic and Syntactic Properties	273
4.2.1.3. Discussion	275
4.2.2. Plesen	276
4.2.2.1. Database	277
4.2.2.2. Variables	281
4.2.2.2.1. Construction Type	281
4.2.2.2.2. Semantic and Syntactic Properties	284
4.2.2.2.3. Discussion	290
4.2.3. Reuen	296
4.2.3.1. Database	297
4.2.3.2. Variables	298
4.2.3.2.1. Construction Type	298
4.2.3.2.2. Semantic and Syntactic Properties	298
4.2.3.2.3. Discussion	299
4.2.4. Longen	300
4.2.4.1. Database	301
4.2.4.2. Variables	302
4.2.4.2.1. Construction Type	302

4.2.4.2.2. Semantic and Syntactic Properties	304
4.2.4.3. Discussion	306
4.2.5. Wanten	309
4.2.5.1. Database	309
4.2.5.2. Variables	312
4.2.5.2.1. Construction Type	312
4.2.5.2.2. Semantic and Syntactic Properties	314
4.2.5.2.3. Discussion	317
4.3. Summary and Global Discussion	320
5. Conclusion	333
Bibliography	338
Primary Sources	338
Secondary Sources	342
Digital Resources	360

List of Tables

Table 1	Word counts of the <i>PPCME2</i> and the <i>PCEEC</i>	129
Table 2	Potential forms of <i>liken</i> in the <i>PPCME2</i> and the <i>PCEEC</i>	139
Table 3	Constituent order types of <i>liken</i> with expletive by frequency	165
Table 4	Period divisions in the <i>PPCME2</i> and the <i>PCEEC</i>	204
Table 5	Distribution of experiencer person for <i>liken</i> in the <i>PPCME2</i> and <i>PCEEC</i>	209
Table 6	Gender of the experiencer argument of <i>liken</i>	213
Table 7	Construction type of <i>liken</i> by experiencer gender	213
Table 8	Overall frequency of (pre-)modal auxiliaries with <i>liken</i>	216
Table 9	Detailed frequency of (pre-)modal auxiliaries with <i>liken</i>	216
Table 10	Logistic regression model of the construction of <i>liken</i>	254
Table 11	Revised logistic regression model of the construction of <i>liken</i>	257
Table 12	Absolute and standardised frequency of human stimuli with <i>liken</i>	262
Table 13	Standardised frequency of <i>liken</i> , <i>plesen</i> and <i>quemen</i> by corpus	278
Table 14	Absolute frequency of <i>liken</i> , <i>plesen</i> and <i>quemen</i> by corpus	279
Table 15	Animacy of the stimulus argument of <i>plesen</i> in full and reduced set	286
Table 16	Absolute and standardised frequency of <i>longen</i> by period	301
Table 17	Absolute and standardised frequency of <i>wanten</i> by period	310
Table 18	Overview of the diachronic development of the construction type	320

List of Figures

Figure 1	Continuum of semantic roles and grammatical relations	103
Figure 1	Absolute and standardised frequency of <i>liken</i> by period	220
Figure 2	Clause type of <i>liken</i> by period	223
Figure 3	Ambiguity of construction of <i>liken</i>	224
Figure 4	Frequency and proportion of the construction of <i>liken</i>	226
Figure 5	Construction of <i>liken</i> without expletive	227
Figure 6	Construction of <i>liken</i> with expletive	228
Figure 7	Animacy of the stimulus argument of <i>liken</i>	229
Figure 9	Abstractness of the stimulus argument of <i>liken</i>	231
Figure 8	Category of the stimulus argument of <i>liken</i>	232
Figure 9	Propositional stimulus arguments of <i>liken</i> without expletive	233
Figure 12	Contrast of human vs. non-human stimuli of <i>liken</i> in full model vs. E2–E3	248
Figure 10	Interaction of propositional vs. non-human stimuli of <i>liken</i> with periods	250
Figure 11	Factor map of chronological, constructional and semantic features of <i>liken</i>	264
Figure 12	Standardised frequency of <i>quemem</i> , <i>liken</i> and <i>plesen</i> per 100,000 words	278
Figure 13	Clause type of <i>plesen</i> by period	280
Figure 14	Frequency and proportion of the construction of <i>plesen</i>	283
Figure 15	Animacy of the stimulus argument of <i>plesen</i>	285
Figure 16	Semantic types of the stimulus argument of <i>plesen</i>	287
Figure 17	Category of the stimulus argument of <i>plesen</i>	288
Figure 18	Semantically propositional stimuli with <i>liken</i> and <i>plesen</i>	292
Figure 19	Frequency and proportion of the construction of <i>longen</i>	303
Figure 23	Semantic type of the stimulus argument of <i>longen</i>	305
Figure 20	Frequency and proportion of the construction of <i>wanten</i>	313
Figure 25	Semantic type of the stimulus argument of <i>wanten</i>	315

Acknowledgements

I would like to thank my supervisor Prof. Dr Volker Gast for giving me the opportunity to engage in a topic that has interested me since the time of my BA studies at University College London and the University of Iceland, for offering continual encouragement and advice on occasions more numerous than I can count, and for his guidance in the execution of the computational methods applied in this study. Without his support, this PhD project would never have developed. I would also like to thank my colleagues at the University of Jena, both former and present, for helpful discussions, constructive feedback, and for creating a friendly working environment. Specifically, I would like to thank Barthe Bloom, Merlijn Breunese, Sally Dixon, Florian Hass, Karsten Schmidtke-Bode, Olena Tykhostup, and Christian Wehmeier, who have all kindly shared their expertise with me. Numerous attendees at our seminars and workshops have commented on early stages of this project, for which I am equally grateful. It is needless to say that I take full responsibility for all shortcomings and errors that remain. The first paragraph and section 1.2 of the introduction, the first part of the theoretical section 2.1 (except sections 2.1.1.2 and 2.1.2.3) and sections 2.2.1.1 and 2.2.2.1.1 from the second part of the theoretical section 2.2 are based on an existing thesis by the present author. The continuum of grammatical relations (Van Valin 2004) and initial ideas regarding the operationalisation of the hypothesis investigated in the present study were equally presented in that work. The discussion of the definition of impersonal constructions as subjectless found in the opening paragraph of section 2.1.1 and in subsection 2.1.1.1 is due for publication in a different revision of the earlier work elsewhere.

1. Introduction

1.1. Impersonal Constructions

Impersonal constructions constitute one of the most fervently discussed topics in diachronic English syntax (for overviews of the topic cf. Visser 1963: 20–35, Mitchell 1985: 427–438, Traugott 1992: 208–213, Fischer 1992: 234–239, Rissanen 1999: 249–252; a very good survey is Denison 1993: 61–102; more recently also Fischer et al. 2017: 145–148). The principal observation about these constructions is their potential lack of an overt nominative constituent. This contrasts with the syntactic rules of Present-Day English (PDE), where non-coordinated finite clauses typically require a nominative constituent, which may be realised by semantically empty constituents like expletive *it*. The fact that such constituents could be absent from comparable constructions in earlier periods of English is illustrated for Old English (OE) by the examples in (1) and (2) below:

(1) **hine** þyrste hwylum and hwylum hingrode

him(ACC-SG) was-thirsty(3-SG) sometimes and sometimes was-hungry(3-SG)

‘he was sometimes thirsty and sometimes hungry’ (Wulfstan, *Hom.* 17, 4)¹

(2) norþan sniwde

north-from snowed(3-SG)

‘it snowed from the north’ (*Exon.* Th. 307, 30)

In (1), the two coordinated verbs *þyrstan* ‘to be thirsty’ and *hyngrian* ‘to be hungry’ occur with a single overt complement, which is marked by accusative case, while the corresponding constituent in the PDE translation is marked by nominative case (also referred to as “subject case”). In (2), the OE verb *sniwan* ‘to snow’ occurs without any verbal complement, but in the PDE translation, semantically empty *it* is supplied in preverbal position, which is the default position of the subject in declarative main clauses and subordinate clauses of Present-Day-

¹ Citations of primary sources are either drawn from the existing secondary literature or from one of the two corpora used in this study. In the former case, the abbreviations of the editions from which the examples are taken are explained in the reference section. In the latter case, the ID code of the relevant corpus is supplied. In order to facilitate the interpretation of the examples, experiencer arguments are generally printed in bold and the forms of impersonal verbs under discussion are underlined, neither of which formatting is present in the original editions or corpora. The quotations from the corpora are otherwise faithful in preserving the additional spacing before punctuation marks, but non-ASCII characters, which are transcribed by combinations of + and the equivalent ASCII characters in the corpora (e.g. *ð* = *+d*) are converted back into their non-ASCII form in order to improve readability. Any grammatical parsing included in the word-by-word glosses is intended to be facilitative and not necessarily exhaustive. The translations are usually the author’s own and hold no claim to literary accomplishment.

English. In either case, the word-for-word equivalent of the OE construction is ungrammatical in Present-Day-English, so there clearly has been a change in the syntactic rules that apply to these verbs. The reason Early-English impersonal verbs and constructions have received special attention by syntacticians and historical linguists alike is at least threefold. First, the variety of syntactic structures commonly discussed under the heading of impersonal constructions, which include cases beyond the ones illustrated above, has led to persistent difficulties and repeated attempts at providing an inclusive but at the same time restrictive definition of the concept. Second, the changes provide a challenge from a comparative point of view, since they have occurred to varying degrees in a range of genetically related languages, including German and the Scandinavian languages. Clearly, one would like to be able to account for the particular development of English compared to that of some of its closest relatives. Third, and most importantly, the changes have been discussed on the basis of different theoretical frameworks of synchronic syntax, whose validity the observed diachronic changes were, in some cases, intended to prove. In terms of their prominence as a topic in diachronic English syntax, impersonal constructions are thus perhaps only second to the emergence of modal auxiliaries as a lexical category of English, which was prominently discussed by Lightfoot in a series of articles and monographs (cf. especially Lightfoot 1979, 1981, 1991 and 1999).

In view of the large amount of existing attention, scholars preceding the present time of writing by no small number of years have felt hard-pressed to motivate their continued engagement with the topic (cf. e.g. von Seeffranz-Montag 1983: 14 or Lagerquist 1985: 123, who calls impersonal constructions a “perennially popular” problem). While the same task would appear to be all the more difficult for the present study, several contributions from more recent years have already demonstrated that academic endeavour can still be rewarded in a subject area that has interested scholars over the course of so many years (cf. Loureiro-Porto 2009, Möhlig-Falke 2012, Miura 2015). Their advancements primarily involve the integration of more recent theoretical concepts derived from functional and cognitive linguistics as well as lexical semantics, and the systematic extension of the empirical basis by the utilisation of electronically compiled corpora, whose increasing availability has brought about something of a paradigmatic shift in the philological sciences. It is in this vein that the present study proposes to contribute to a longstanding discussion by subjecting hypotheses that were formed in earlier theoretical treatments on the basis of more limited data to the scrutiny of a larger corpus study that approaches the topic on the basis of quantifiable evidence. In contrast to the aforementioned studies, its focus will be on the development of impersonal verbs during Middle English (ME) and Early-Modern English (EModE), since these periods constitute the phase during which

most of the relevant changes are observable. A more detailed outline of the present study will be given at the end of this introductory section. Its distinction from previous contributions will be more readily appreciated after a brief survey of some of the earlier literature, which cannot be exhaustive given the large number of entries in the field. While the emphasis of this investigation will be on a specific subset of impersonal verbs, the study will not attempt to bypass the customary discussion of the definitional problems of impersonal verbs in general, which will serve to both define the object of interest more narrowly and to highlight the broader scope of impersonal constructions as a syntactic phenomenon. The definition will take place in the first part of the theoretical section, which also provides a more detailed account of earlier classifications and theoretical treatments of impersonal verbs. The introductory paragraph of this introduction and the subsequent section 1.2 as well as most of the theoretical section 2.1 and the accounts of reanalysis (section 2.2.1.1) and actualisation (section 2.2.2.1.1) in section 2.2 are based on an existing thesis by the present author. Since these parts are fundamental to the understanding of the theoretical background and the methodological design of the empirical investigation presented in the later sections of this study, they are repeated here in a slightly revised and considerably expanded form, but remain essentially unchanged.

1.2. Survey of Earlier Literature

Any survey of earlier literature on the topic of impersonal verbs will necessarily have to remain eclectic due to the large number of existing contributions, but among all earlier treatments of note, van der Gaaf (1904) is most commonly cited as the pioneering study of English impersonal verbs. Building on earlier observations by Jespersen (1894), the author provides a discussion and extensive exemplification of impersonal verbs and their major developments during the history of English. For the OE period, he identifies about 40 impersonal verbs, which include constructions involving predicative adjectives like *laþ beon* ‘to be odious, hateful’ and *leof beon* ‘to be dear, agreeable’. Although their construction with a dative experiencer argument is essentially equivalent to that of other impersonal verbs, such “phrasal impersonals” have been mostly disregarded in the more recent literature (an exception is van der Wurff 1992). The exact number of lexical items in impersonal constructions also depends on whether compound verbs such as OE *ofþyncan* ‘to cause regret, displeasure, offence’, *misþyncan* ‘to give a wrong idea’ and *geþyncan* ‘to seem, appear’ are regarded as separate items besides the base verb *þyncan* ‘to seem’. A range of additional verbs enter the English language during the ME period, cf. e.g. ME *dremen* ‘to dream’, a loan from Old Norse, cf. ON *dreyma* ‘id.’, which replaces OE *mætan*

'id.', and *greven* 'to cause grief', a loan from Old French, cf. OF *grever* 'to irritate, bother'. The list of impersonal verbs in Visser's *Historical Syntax* (cf. Visser 1963: 20–23), which combines items from Old English and Middle English, contains about 80 different verbs with an oblique experiencer argument like *þyrstan* and *hyngrian* in (1) above, while zero-argument verbs like *sniwan* in (2) above amount to about 30 instances (cf. Visser 1963: 36f.). Taken together, the historical records thus comprise at least 100 verbs that occur in some form of impersonal construction during at least one of the periods of Early English.

The first study that attempted a synchronic classification of impersonal verbs in Old English was conducted by Wahlén (1925), who makes a rather broad structural distinction between verbs in simple sentences and verbs in complex sentences containing an embedded complement clause. Of greater interest is his rarely cited distinction between different semantic classes, including natural phenomena, physical and mental affections, and the course of events, which immediately raises the question of the conceptual basis of impersonal constructions. A more elaborate set of structural subclasses and a slightly less comprehensive set of semantic classes were developed by Elmer (1981), whose diachronic account essentially follows van der Gaaf in recognising three major developments of impersonal verbs, viz. lexical obsolescence, personalisation, and the introduction of an expletive. The obsolescence of OE impersonal verbs is already observable during the Early-Middle-English (EME) period between 1150–1300 (cf. van der Gaaf 1904: 12f.), while new impersonal verbs emerged as a result of borrowing or the reinterpretation of indigenous verbs. From around 1300 onwards, a gradual personalisation of impersonal constructions took place by the introduction of nominative experiencer arguments, which frequently coexisted alongside impersonal constructions until about 1500 (cf. van der Gaaf 1904: 142). The third development constitutes the introduction of an expletive pronoun into preverbal position, which frequently co-occurs with a PP experiencer, e.g. *it seems to me* and *it behoves us* (cf. van der Gaaf 1904: 36–39). These remain a viable option also in Present-Day-English. It should also be noted that expressions such as *methinks* or *meseems* were relatively widespread during the EModE period, although they probably need to be regarded as lexicalised constituents and rather than instances of productive use of impersonal constructions (cf. Rissanen 1999: 251). In addition, Wahlén's type involving complex sentences appears to have been particularly resilient, continuing with certain verbs until the 19th century (cf. Elmer 1981: 159). Also Visser (1963: 35) notes the occurrence of oblique experiencer arguments as late as the 16th-century writings of Thomas More. Overall, the general picture that emerges from these earlier studies is that the transition from impersonal to personal constructions did not proceed in a homogeneous fashion.

The different structural and semantic properties of verbs entering into impersonal constructions and the different diachronic tendencies observable among these verbs clearly make any systematic account of their development a complex issue. The largely descriptive nature of van der Gaaf's, Wahlén's and Elmer's work leaves room for further theoretical treatment, and a number of proposals formalised in various theoretical frameworks have been put forward since then. The most influential contribution in stimulating theoretical debate was probably Lightfoot's treatment of impersonal verbs in his *Principles of Diachronic Syntax* (1979), an account grounded in the Extended Standard Theory of transformational grammar. In this treatment, Lightfoot proposes a diachronic mechanism of reanalysis, whose essence can be traced back to the discussion of morphological surface ambiguity in Jespersen 1927. The account will be presented in some more detail in the theoretical section below (cf. section 2.2.1.1), but it should already be noted at this point that various criticisms were launched against the theoretical notion of reanalysis soon after its reinstatement by Lightfoot. One of these concerns a perceived neglect of the empirical evidence. According to an influential proposal by Fischer/van der Leek (1983) and a later revision (1987), the evidence points to a parallel occurrence rather than a succession of impersonal and personal constructions. Based on the theory of Government and Binding, this co-occurrence is accounted for by optional case assignment in the relevant lexical frames, and the diachronic development is interpreted as an instance of lexical rather than syntactic reanalysis. The notion of optional case assignment was challenged by Allen in several articles (1986a, 1986b, cf. also 1997), culminating in an extensive monograph (1995) in which Allen put forward her own account of the development of impersonal verbs and impersonal passives within the framework of Lexical Functional Grammar. Essentially, she regards the development as a replacement of lexical case marking by structural case marking, but only in the case of passives does she consider this replacement to be the more or less immediate result of case syncretism, while the subjects of impersonal verbs are described in terms of a more gradual alignment with the general tendency towards structural case marking. Allen's assumption of such a structural tendency is reminiscent of the notion of typological drift invoked by von Steffens-Montag in her comparative study of impersonal verbs (1983), although neither assumption appears to provide much of an explanation of how the specific change unfolds. A particular merit of Allen's work, however, is her dedication to philological scrutiny and her consistent reference to empirical evidence in the form of concordances and a personal compilation of primary texts. It should also be noted that Lightfoot's account of reanalysis was subjected to several revisions, which ultimately aligned it with the theory of Principles and Parameters (Lightfoot 1991, 1999) and incorporated

the notion of speaker-internal diglossia in order to account for the apparent gradualness of change and the observed synchronic variation between different constructions. The most recent discussion of this proposal to be considered in some detail below is Roberts 2007, whose view of parametric change as the pervasive mechanism of diachronic syntax is formulated in a more current version of transformational grammar according to the Minimalist Program, although his account of impersonal verbs is much beholden to Allen 1995.

Besides the aforementioned studies, several dissertations from more recent years were devoted to the study of impersonal constructions. Of these, Loureiro-Porto 2009 is a true testimony to the assiduous work required of scholars conducting lemma-based corpus studies on historical corpora that lack lemmatisation, a problem which will be discussed further below in the methodological section (cf. sections 3.1 and 3.2). Loureiro-Porto's compilation of several historical corpora into a single diachronic database serves as a model for the present investigation, although her discussion of the circumscribed lexical field of semantic predecessors of PDE *need* and their interpretation in terms of external vs. internal deontic modality largely remain outside of the focus adopted here. The second study, which equally involves a large historical database, is Möhlig-Falke 2012. Her intention is to complement earlier formal accounts of impersonal verbs with a functional perspective, drawing on theoretical concepts ranging from semantics and discourse-pragmatics to cognitive linguistics and construction grammar (cf. Möhlig-Falke 2012: 22–25, 51). The broad design of her study, which is aimed at complete coverage of impersonal verbs during the OE period, makes it understandable that her use of corpora remains limited to Old English, although an outlook on the later ME and EModE periods is provided with the help of the *Middle English Dictionary (MED)* and the *Oxford English Dictionary (OED)*. The corpus study itself is based on the *Dictionary of Old English Corpus (DOEC)*, which constitutes the most comprehensive database of Old English at the time of writing, although it is not a balanced corpus in the strict sense of the word. Möhlig-Falke uses these data to quantify the different syntactic complementation patterns of impersonal verbs within each of the semantic groups that she distinguishes, thus aiming to determine the overarching function of impersonal constructions and the way in which lexical and constructional meaning interact in OE impersonal verbs. It is important to note, however, that her definitional criteria result in a different distinction of impersonal and personal constructions compared to the one applied in the present study (cf. section 2.1). Specifically, she regards constructions involving a nominative constituent or an expletive pronoun as personal constructions, even if the experiencer argument continues to be marked by oblique case. This obviously affects the proportions of impersonal and personal constructions

established by Möhlig-Falke (2012: 115f.), and it means that a verb like OE *lician* ‘to please’ and its compounds, which she regards as impersonal verbs, are, in fact, considered to be constructed personally in over 75% of all attested instances. This discrepancy does not seem to affect Möhlig-Falke’s overall characterisation of impersonal constructions as conceptually backgrounding the initiator of a dynamic process, however, even though this may be equally claimed about the instances that she excludes from the definition. Her attribution of the apparent variation between impersonal constructions and transitive constructions to different degrees of intentional involvement or responsibility of the experiencer argument, on the other hand, will be important in later parts of the present study, although it will be considered more systematically for constructions during the ME and EModE periods, for which Möhlig-Falke only provides a preliminary outlook.

The most recent dissertational study to be considered in the present context is Miura’s treatment of impersonal verbs of emotion (Miura 2015), which differs from Möhlig-Falke’s study in being explicitly devoted to the constructions of Middle English. The author’s primary goal is to supplement earlier theoretical accounts with a perspective on lexical semantics, by which she intends to account for the difference between verbs of emotion that occur in impersonal constructions and verbs of emotion that appear to be exclusively used in personal constructions. This proves to be a difficult task, since common verbs of emotion such as *love* and *hate* are notably absent from impersonal constructions during all historical periods (cf. also Krzyszpień 1990: 77 with a similar remark). The main resources employed in her study are the *Historical Thesaurus of the Oxford English Dictionary (HTOED)* and several historical dictionaries, most prominently the *MED*. A major advantage of the dictionaries is that they contain lemmatised entries of the verbs under investigation and thus offer immediate access. Miura makes it clear, however, that they are not designed with the requirements of a quantitative corpus study in mind, since the illustrative citations included in them, although quite extensive in many cases, do not yield reliable frequency counts of the different construction types. She consequently characterises her investigation as a predominantly qualitative study with a particular focus on the verbal semantics rather than the quantitative data of different verbs. It is also important to note, once more, that the emphasis of her account lies on the semantic difference between verbs that exhibit (any number of) impersonal constructions and verbs that never occur in impersonal constructions, not on the synchronic difference between impersonal and personal constructions of a given verb or its diachronic development. According to Miura (2015: 9), an account of the latter difference would require an extensive corpus study beyond the period of Middle English.

1.3. Motivation and Outline of the Present Study

Although many important contributions to the ongoing discussion of impersonal verbs and their diachronic development in English were left unnamed or mentioned only in passing above, the brief survey already points in the direction that the present study is going to take. The discernible focus on different formalisations of syntax in earlier discussions of the diachronic development of an apparently disparate set of lexical items has been complemented more recently by studies inspired by usage-based frameworks, which place greater emphasis on the role of semantics, discourse pragmatics and the communicative function of constructional variants. The notion that syntactic patterns are not simply autonomous but constitute discrete form-meaning pairings of their own naturally suggests that semantic factors may have played a role in the diachronic development of impersonal verbs as well. The basic idea is already expressed by Allen, who discusses several instances in her data in terms of semantic differences between different constructions and different lexical items (cf. Allen 1995: 326–347). Allen states, however, that “[a] thorough investigation of the semantic factors involved in the changes which took place in the experiencer verbs would be well beyond the scope of the present study, and would in fact easily fill a whole book on its own” (ibid.: 326). In a similar vein, Möhlig-Falke concludes that the synchronic variation between nominative and oblique experiencer arguments in Old English can be related to different semantic interpretations of the experiencer argument (cf. Möhlig-Falke 2012: 231), and she goes on to suggest that a “semantic and discourse-pragmatic analysis of impersonal expressions might fruitfully be extended to the period of 1150 to 1650 in a corpus-based investigation” (Möhlig-Falke 2012: 235), which would offer a larger and more easily quantifiable database than the historical dictionaries used in her outlook on the diachronic development of impersonal verbs.

While a large-scale investigation of impersonal verbs during the ME period is already available in the form of Miura’s study, this study is explicitly not corpus-based and has a different focus than the present one. In motivating her choice of resources, Miura refers to various shortcomings of the available corpora in terms of size and their representation of different text types, and she ultimately decides to use the citations in the *MED* as her primary database, which she also considers to be better suited to the task of investigating the lexical semantics of impersonal verbs of emotion (cf. Miura 2015: 96f.). Another recent study involving impersonal verbs in Middle English is Trips/Stein 2019 (cf. also the report in Percillier 2016), which explores the notion of contact-induced change in a range of ME impersonal (and other) verbs based on the model of Old French. Their discussion involves some of the same verbs treated in the present study and includes an appeal to Allen’s assumption of

semantic factors in the process of change, but their research focusses specifically on the replacement of NP constituents with PP constituents and not on the general change from impersonal to personal constructions. It appears, then, that no full-fledged corpus study of potential semantic factors regarding the development of impersonal constructions with oblique experiencer arguments into personal constructions with nominative experiencer arguments has yet been conducted. The execution of such an investigation will hence be the primary goal of the present study. While it does not constitute a pioneering study in the field of historical linguistics, the implementation of corpus resources in the investigation of historical data nevertheless requires a number of individual solutions to different methodological problems, some of which are already hinted at in Miura's reservations about the use of ME corpora. The present study also differs from existing ones by attempting to provide a model of diachronic change using both descriptive and inferential statistics. In doing so, it will primarily strive to achieve methodological advances rather than to apply a novel theoretical framework to old data. Instead, the theoretical discussion is rooted in the generative discourse that has dominated much of the earlier work on impersonal verbs and diachronic syntax in general. The present investigation of semantic factors in the development of impersonal verbs does not, however, seek to establish the validity of any particular version of generative theory, but, instead, adopts a usage-based view of language that regards formal expressions as intrinsically tied up with meaning.

The organisation of the remainder of the text is as follows. Section 2 provides the theoretical background to the investigation. It begins with the definition of impersonal verbs and considers some of the problems associated with the task of finding a common denominator for all constructions commonly discussed under this heading. It also provides an overview of some of the major semantic and syntactic classifications that have been applied to impersonal verbs. It then proceeds with a more detailed discussion of earlier diachronic accounts, which lead up to the present research hypothesis. The starting point is the traditional reanalysis account that was proposed in earlier generativist frameworks. This theory of reanalysis is complemented by the notion of a more gradual actualisation of change, whose application to the diachronic development of impersonal verbs will be discussed in the final part of section 2. It should be noted, however, that the present study is not intended as proof of a generativist view of language change, but, instead, conforms more readily with a theoretical conceptualisation of language as a set of emergent and changeable structures, in which syntactic and semantic components are intrinsically intertwined. Another central concept in the traditional theory is the notion of markedness, which is viewed as an important constraint on the directionality of change. Since

the specific change under investigation is hypothesised to involve an effect of semantic factors, the application of the theory of markedness to the development of impersonal verbs will entail a discussion of semantic roles and their hierarchical mapping onto grammatical relations. Drawing on the definition of prototypical agent features, the hypothesis will be specified in terms of a chronologically and quantitatively privileged spread of nominative experiencer arguments to contexts that are susceptible to a more agentive interpretation of the experiencer argument. The operationalisation of agentive properties for the purposes of the corpus study will conclude the theoretical section.

Section 3 discusses the methodological considerations and the general procedure of the present study. Following a number of general caveats about the nature of the primary sources, the selection of the corpora that form the primary database will be motivated. Since these corpora are not lemmatised like the historical dictionaries used in most of the earlier studies discussed above, the process by which the relevant verbal tokens were extracted from the corpora will be described. This will be followed by a discussion of the necessary restrictions of the dataset, which involve the exclusion of non-finite, passive and ambiguous constructions. The main part of the methodological section will then be dedicated to the process of data annotation, both with regard to the criteria used to identify the construction type and with regard to the semantic and syntactic properties that constitute potential predictors. Since the process of annotation is potentially susceptible to a certain degree of subjectivity, which would be masked by seemingly objective findings of a later quantitative evaluation, the aim will be to ensure as much transparency about individual analyses as can be justified within the applicable confinements of space. At the same time, the section will serve as an illustration of the source materials by providing citations of representative examples of all categorial distinctions that were made. This will serve as philological basis for the more abstract process of representing the observed diachronic changes with the help of a logistic regression model, the central elements of which will be outlined at the end of the section.

Section 4 will present the results that were obtained from the annotated datasets of the verbs under investigation. Following a descriptive overview of the diachronic distribution of the different variable levels, the configuration of the regression model in terms of periodisation of the observed changes, a refinement of the conceptual variables, and in terms of the selection of variables to be used in the final model alongside potential variants will be discussed. As such, the results section continues to develop some of the theoretical concepts discussed in earlier parts the study and tests them against the tendencies that emerge from the data. The interpretation and discussion of the results will begin with the findings of the logistic regression

model, which essentially constitutes an expression of the associations observed in the data, but it will also take any additional observations into account that emerge from a qualitative investigation of the attested instances of the relevant verbs, especially since not all datasets qualify for logistic regression. Finally, a global discussion of the results obtained individually for each verb will take place and compare the findings with the predictions of the hypothesis formulated in the theoretical section. The discussion will be followed by a general conclusion and an overall evaluation of the study in section 5.

2. Theoretical Background

2.1. Impersonal Verbs

2.1.1. Definition

The inherent difficulty of defining impersonal verbs and constructions is frequently remarked upon by scholars working in the field. Visser (1963: 20) considers the traditional terminology a misnomer, and Méndez-Naya/López-Couso (1997) devote a whole article to the question of what “impersonal” really means. Problems arise not only from the fact that different scholars employ different terminological systems, which include distinctions like “really impersonal” and “quasi-impersonal” (van der Gaaf 1904: 2), “true impersonal” (Anderson 1986: 168) and “semi-impersonal” (Anderson 1988: 8) or, on a more subtle level, between “impersonal” with inverted commas and “impersonal” without (Ogura 1986: 16). Nevertheless, it would be unjustified to consider the different terminological distinctions a mere “terminological maze created by previous scholarship” (Möhlig-Falke 2012: 12), since they are the result of the genuine complexity and diversity of the linguistic constructions under discussion. At the same time, the term “impersonal” itself is ambiguous and open to several interpretations. In its most literal sense, the term can be taken to refer to the absence of persons or participants involved in a given situation (cf. Denison 1993: 62). This would presumably limit its scope to verbs such as *sniwan* cited in (2) above and comparable references to atmospheric conditions or to the time of day (cf. e.g. ME *nighten* ‘to grow dark’), which do not normally involve any participants. Verbs such as *pyrstan* and *hyngrían* cited in (1), on the other hand, require an animate participant that experiences the sensation described by the predicate, and they thus require a different interpretation of the term “impersonal”. According to Fischer/van der Leek (1983: 347), such experiencer arguments constitute one of the defining features of impersonal constructions, and also Allen (1995:1–2) specifically uses the term “experiencer verbs” in her

investigation of impersonal verbs and constructions. A definition based solely on the presence of an experiencer argument excludes zero-argument verbs like *sniwan*, however, which other scholars regard as prototypical examples of impersonal constructions, while it does not rule out constructions with a nominative constituent, which many scholars consider to be personal. Depending on the particular focus, then, the notion of what “impersonal” really means can vary, but verbal argument structure and semantic roles such as that of experiencer clearly play a role in the definition.

2.1.1.1. Subjectless Constructions

The most frequently cited definition of impersonal constructions focuses on their syntactic features rather than argument structure or semantic roles. These features comprise (a) the absence of a nominative constituent and (b) the lack of verb agreement (cf. e.g. Denison 1993: 61, Traugott 1992: 208, Anderson 1988: 8, with additional observations on case marking of the expressed arguments Möhlig-Falke 2012: 6, followed by Miura 2015: 4, cf. also McCawley 1976: 192). In accordance with the second criterion, the term “impersonal” can thus be understood as a reference to the lack of person-number variation in the verb, which, instead, occurs in the third-person singular by default (for the tradition of this interpretation cf. Jonson 1640 cited by Pocheptsov 1997: 482³ and Möhlig-Falke 2012: 6 with reference to Sweet 1891: 93). A definition along these lines neatly subsumes one-argument experiencer verbs like *hyngrian* and *þyrstan* in (1) and weather verbs like *sniwan* in (2), irrespective of the presence or absence of additional participants. Furthermore, since control of verb agreement and nominative case marking are typical subject properties of Present-Day-English, the alternative term “subjectless” can be applied and is even preferred by some scholars as definition of impersonal constructions (cf. e.g. Anderson 1986 and von See Franz-Montag 1983: 13).

The definition of impersonal constructions as subjectless is fraught with problems of its own, however. The main problem is that literally all types of complements that occur besides the verb in impersonal constructions can and have been considered subjects under different analyses. These complements are illustrated in (3)–(6) below:

(3) hit rine and sniwe and styrme ute

it rain(3-SG-SUBJ) and snow(3-SG-SUBJ) and storm(3-SG-SUBJ) outside

‘it may rain and snow and storm outside’ (Alfred, *Bede* 2, 13)

- (4) **him** eglde þæt he was betra þonne he
 him(DAT-SG) ailed(3-SG) that he was better than he
 ‘it ailed him that he was better than him’ (Alfred, *C. P.* 234, 8)
- (5) **us** sceamaþ to secgenne ealla þa sceandlican wiglunga
 us(ACC-PL) shames(3-SG) to mention all the disgraceful auguries
 ‘we are ashamed to mention all the disgraceful auguries’ (Ælfric, *Hom.* 1, 370, 100)
- (6) **þam wife** þa word wel licodon
 the lady(DAT-SG) these words(NOM-PL) well pleased(3-PL)
 ‘the lady liked these words well’ (*Beow.* 639)

The example in (3) is similar to the one (2) above in that it contains three instances of verbs that refer to weather conditions without any explicit participants. It differs from the earlier example, however, in that it also contains a semantically empty pronoun *hit* ‘it’, which occurs in preverbal position. The examples in (4) and (5) each contain a preverbal experiencer argument (*him* and *us*) and are thus similar to the example in (1) above, but, in addition, they contain either a finite complement clause (*þæt he was betra þonne he* ‘that he was better than him’) or a non-finite complement clause (*to secgenne ealla þa sceandlican wiglunga* ‘to mention all the disgraceful auguries’). The example in (6) contains an oblique experiencer argument *þam wife* ‘the lady’ and an additional nominal complement *þa word* ‘these words’. The potential subject status of the additional constituents in (3)–(6) and the question of whether all of the constructions above can be subsumed under a uniform analysis will be discussed in the following paragraphs.

2.1.1.1.1. Expletive *Hit*

Various answers to the question raised above have been given with regard to expletive *hit*, which neatly pinpoint the conflict that arises from the application of either semantic or syntactic criteria to the definition of impersonal verbs and constructions. Wahlén (1925: 8–11), for example, argues with reference to Brugmann (1917: 2) that *hit* in constructions such as (3) has “no real semantic function” (p. 8), and he therefore considers expressions without *hit* “no less impersonal in character than those containing that word” (p. 10). His view is shared by Mitchell (1985: 429–31), who relies heavily on Wahlén on this point. Mitchell defines the impersonal

construction as “one which has only the formal subject *hit* [...] or which has no expressed subject and for which no subject other than the formal *hit* can be supplied” (p. 427). Essentially, this is a semantic definition, based on the absence of a participant identifiable as agent or force, and the optional presence of a semantically empty “formal subject” *hit* is merely considered to be an expression of this absence. Also van der Gaaf (1904: 2), while restricting his definition of “really impersonal” verbs to those that “express natural phenomena [...] and can have no other subject than *it*”, does not make a general distinction between instances with or without an expletive, neither in the case of “really impersonal” weather verbs nor in the case of “quasi-impersonal” experiencer verbs. His terminological distinction is rather based on the presence or absence of a personal argument (cf. similarly Pocheptsov 1997: 470), besides which the presence of *hit*, again, constitutes an optional epiphenomenon. The alternative view is held by Anderson (1988: 8, 1986: 168), who draws the line between “true impersonal” and “quasi-” or “semi-impersonal” constructions according to whether or not expletive *hit* is present (cf. also Kim 1999: 335, who points to positional parallels of *hit* with nominative NPs to motivate this distinction). The main argument in favour of Anderson’s view is that the presence of *hit* violates the definitional criterion (a): in structural terms, it constitutes a nominative subject constituent, which merely lacks semantic content. If impersonal constructions are, indeed, defined as subjectless, this implies that constructions containing expletive *hit* are not impersonal. While the observation of a structural difference between constructions with and without an expletive may be accurate, the theoretical implication of terms such as “quasi-impersonal” or “semi-impersonal” remains unclear, however, and their coinage appears to be an acknowledgement of the fact that it seems unfortunate to separate examples such (2) and (3) despite the fact that they essentially convey the same semantic content.

The fact that *hit* constitutes a semantically empty constituent in (3) also distinguishes it from other cases of impersonality that have traditionally been considered under this heading (cf. e.g. Goebel 1888: 21, Brandenstein 1928: 9). The latter type can be observed in sentences containing the German expletive *es*, which has a wider scope than PDE *it* and extends to impersonal passives of the type *es wurde getanzt und gelacht* ‘there was dancing and laughter’ and sentences with indefinite subjects like *es klopft* ‘someone is knocking’. The means of expressing this kind of pragmatic impersonality in Present-Day-English include existential *there* or indefinite *someone* besides many others, but such cases are clearly distinct from the type of impersonality that is at issue here, since they presuppose the existence of an agent or force, which is either not expressed or not identifiable in the given context. As such, they are ruled out from the definition of “impersonal” according to Mitchell’s formulation cited above.

Natural events like the precipitation of *rain* or *snow*, on the other hand, while obviously resulting from specific conditions like humidity and temperature, are generally not expressed with reference to such forces by the average speaker. Due to the inherent difficulty of identifying the agent or force behind these phenomena, weather verbs have thus been considered prototypical examples of impersonal verbs. Similar observations can be made about physiological states like *hunger* and *thirst*. It is not easy, for the average speaker, to identify the cause or source of these sensations, which therefore typically remain unexpressed, unless the verbs are used figuratively in the sense of longing. For physiological sensations such as pain, which is referred to in impersonal constructions of ME *smerten* ‘to smart, cause pain’, a body part may be expressed as the formal source, even though it would probably be conceived of more accurately as the location rather than the source of pain. Nevertheless, the semantic definition of impersonal constructions as those “for which no subject other than the formal *hit* can be supplied” seems to accurately distinguish between the two types of impersonality discussed in this paragraph, and the optional presence of an expletive does not affect the cognitive basis which appears to be underlying the phenomenon of impersonal constructions. Whether the two constructions in (2) and (3) require different analyses on syntactic grounds has not been answered so far, however.

An apparent difference between the two examples is the presence of the locational adjunct *norþan* in preverbal position in (2). One could assume that the fronting of this element accounts for the absence of expletive *hit*, and that expletive *hit* would otherwise surface if no constituent were available for fronting or simply not fronted depending on the organisation of information structure. According to this assumption, the syntactic rules underlying the two examples would be identical, and (2) would be no more impersonal than (3), since one would expect *hit* to occur also in (2) if *norþan* were not fronted. An explanation along these lines would also accord with the observation that the presence of *hit* is, in fact, preferred with expressions of natural phenomena (cf. Mitchell 1985: 430), since their typical lack of semantic arguments leaves fewer candidates for fronting. Counterexamples to this explanation do exist, however, as illustrated by the example in (7):

(7) þa rinde hit

then rained(3-SG) it

‘then it rained’ (OE *Gosp. Mt. 7, 27*)

The occurrence of *hit* in conjunction with a fronted constituent *þa* ‘then’ in this example could, of course, be an extension of a formerly more circumscribed distribution, but, as it stands, the example demonstrates that the presence of a fronted element does not restrict the occurrence of *hit* and that its absence in (2) is, in fact, contrastive.

A closer examination of the distributional properties of *hit* in Old English was conducted by Allen (1986b: 466f.), who investigated constructions containing a clausal complement like the one in *him eglde þæt he was betra þonne he* ‘it ailed him that he was better than him’, cited in (4) above, to which one can add (8), Allen’s (4):

(8) *þa gelamp hit þæt æt þam gyftum [...]*

then happened(3-SG) it that at that wedding

‘then it happened that at that wedding’ (Ælfric, *Hom. Th.* 1, 58, 12)

While (4) contains a preverbal experiencer argument *him* and no instance of *hit*, example (8) demonstrates, once more, that the occurrence of the expletive is not generally restricted by the presence of a fronted constituent. Allen’s overall findings with regard to the constituent order in constructions of OE *gelimpan* ‘to happen’ are that *hit* occurs without a fronted element (*hit-V*) in 29% of all 198 instances in prose declarative main clauses without an experiencer argument, while it occurs in postverbal position alongside a fronted element (*X-V-hit*) in almost 38% of all instances and in preverbal position alongside such a constituent (*X-hit-V*) in almost 33% of all instances. These findings suggest that the occurrence of *hit* alongside a fronted element was, in fact, rather frequent, and they weaken an earlier proposal by Elmer (1981: 56), who traces the distribution of *hit* to a V2 target in Old English, since V2 is actually prevented in the last-mentioned case, if pronominal *hit* is not simply analysed as clitic. There is a notable difference between these instances of *gelimpan*, however, and those that contain an experiencer argument. Although the evidence is less copious in this case, the absence of *hit* in all 17 examples from main clauses and in all but 1 of 11 subordinate clauses (cf. Allen 1986b: 468) strongly suggests that the distribution of *hit* was at least sensitive to the presence of an experiencer argument. Its frequent occurrence with *gelimpan* can thus be related to the fact that the experiencer argument is not an obligatory element in the semantic frame of verbs expressing fortuitous events, and the distribution of *hit* consequently appears to align with the argument structure of the relevant verbs rather than with the presence or absence of fronted elements in general. The introduction of a “formal subject” *hit* may consequently serve to distinguish the relevant instances from impersonal use defined in terms of lack of a subject, but, for the OE

period at least, this is largely equivalent to restricting the definition to verbs accompanied by an experiencer argument, since weather verbs and one-argument instances of verbs of happenstance most frequently occur with expletive *hit*.

2.1.1.1.2. Clausal Complements

Similar considerations are necessary with regard to finite complement clauses like the ones in (4) and (8) and non-finite complement clauses like the one in (5). The two former examples illustrate two different cases: in (8), the clausal complement is preceded by expletive *hit*, which can be analysed as the formal subject of the clause in accordance with the discussion above, while no such constituent is present in (4), and the clausal complement itself needs to be considered a potential candidate for subjecthood. Since clausal constituents are not marked for case, however, one of the major properties of PDE subjects, at least, is unavailable. An alternative option would be to infer the subject status of clausal constituents from their co-occurrence with an expletive pronoun: since *hit* can be interpreted as being covertly marked by nominative case, a subject interpretation seems to be available for clausal complements as well. There is no agreement, however, as to whether such an analysis is permissible. Traugott (1992: 217) rejects the subject interpretation of clausal complements based on the perceived lack of evidence from semantic criteria and word order. In fact, most scholars regard clausal complements in the context of impersonal constructions as oblique constituents. Curiously, evidence for this is gathered from their co-occurrence with oblique pronouns as in (9):

(9) þæs **us** scamaþ swyþe þæt we bote aginnan

that(GEN-SG) us(ACC-SG) shames(3-SG) very

‘we are very ashamed to begin atonement’ (Wulfstan, *Hom.* 165, 10 fn.)

In this example, the genitive pronoun *þæs* is said to have cataphoric reference to the subsequent clause, which, consequently, cannot be analysed as subject (cf. Traugott 1992: 236, Fischer/van der Leek 1983: 349). Yet it is difficult to see how such examples differ from the one in (8), where nominative *hit* would suggest a subject interpretation, other than that demonstrative *þæs* might be considered slightly more referential than anaphoric *hit*. It is also difficult, then, to decide on the basis of case marking which analysis should be extended to clausal constituents in examples like (4), where no additional pronoun is present.

Another option would be to analyse clausal constituents as subjects on the basis of verb agreement. The grammatical properties of clausal constituents are probably best defined as third-person singular, which means that they are consistent with the morphological form of the verb in the examples above. Agreement features are no help in identifying clausal complements as subjects, however, since the verb would be expected to inflect for third-person singular by default also in subjectless constructions. Distributional properties are therefore commonly referred to as the decisive criterion for establishing the syntactic status of clausal constituents. Anderson (1988: 11, 1986: 173), for example, rejects the subject analysis of clausal constituents on the basis of their consistent occurrence in postverbal position in Old English, while preverbal position is considered to be the unmarked position of the subject (cf. similarly von Seeffranz-Montag 1983: 45). There appears to be some inconsistency in Anderson's simultaneous claim (1988: 9) that the frequent occurrence of oblique experiencer arguments in preverbal position does not constitute evidence of their potential subject status, but it is true, in any case, that clausal constituents are notably absent from preverbal position in Old English. Also Visser (1963: 19, 25f.) dismisses the subject analysis of finite complement clauses with reference to their distributional properties, but he is less firm on the matter with infinitival complements like *to secgenne [...]* in (5) above, since these occasionally also occur in preverbal position. It seems an obvious point, however, that the relative heaviness of clausal constituents in terms of phonological and syntactic weight is a potential factor in their failure to occur in preverbal position. As such, they could be considered subject constituents with an additionally motivated restriction on their syntactic distribution. Such an analysis is advocated by Mitchell (1985: 433), but only for instances with expletive *hit*, while instances without an expletive are left undecided. Since neither case marking, verb agreement or distributional properties appear to constitute fully satisfying criteria for the disambiguation of the syntactic status of clausal constituents in Old English, the matter will be left undecided here as well, although the analysis favoured for Present-Day-English suggests that clausal constituents can be subjects. A similarly noncommittal approach is adopted by Denison (1993: 64f.), who interprets all instances of clausal complements as ambiguous apart from those with a provisional genitive pronoun. It should be noted, however, that such an approach implies the potential of significantly reduced scope of impersonal constructions defined as subjectless constructions, since finite and non-finite clauses are very common as complements of impersonal verbs like *gelimpan* as well as a range of other verbs that do not occur with genitive complements and therefore remain ambiguous in terms of the presence of a subject.

2.1.1.1.3. Nominative Complements

The last type of complement that potentially conflicts with the definition of impersonal constructions as subjectless involves nominal complements like *þa word* ‘these words’ in (6) above. Such constituents, unlike clausal complements, can be overtly marked by nominative case and thus require an analysis as subjects according to most scholars (for differing views with regard to Old English cf. von Seeffranz-Montag 1983: 45 and the discussion of Allen 1986a below). The position of these constituents can be preverbal as in (6), but postverbal position equally occurs, cf. (10) below:

(10) **þe licab** se almihtiga god

you(ACC-SG) pleases(3-SG) the almighty God(NOM-SG)

‘the almighty God pleases you’ (*Shrn.* 196, 35)

Visser (1963: 24f.) considers the status of such nominal complements ambiguous when they are marked by “zero-case”, a term by which he refers to the indistinct nominative-accusative case marking that is typical of many nominal declensions in Old English. This type of case marking generally apply to neuters such as *þa word* in (6), while instances like *se almihtiga god* ‘the almighty God’ in (10) are unambiguously marked for nominative case on the demonstrative determiner and the adjectival form. Regarding the status of ambiguously marked nouns, Visser perceives a continual increase in their occurrence in preverbal position, which suggests a subject analysis for the period postdating Old English at least. Most scholars regard nominative case marking of the unambiguous examples as sufficient evidence of their general status, however, and analyse the relevant constituents in (6) and (10) as subjects regardless of their position (cf. e.g. Fischer/van der Leek 1983: 350). The analysis is supported by the apparent control of verb agreement on the plural form *licodon* ‘pleased’ in (6), which indexes the grammatical properties of *þa word*. Taken together, nominative case marking and control of verb agreement are strong indications of subjecthood, which seem to require the exclusion of the relevant instances from a discussion of impersonal constructions under the heading of subjectless constructions, and, indeed, most scholars regard them as personal.

Despite this fact, instances with an additional nominative complement are frequently discussed in the context of impersonal constructions. The main reason seems to be that the relevant instances are formed from the same kind of verbs that enter into constructions regarded as impersonal due to the inherent morphological ambiguity and distributional restrictions of clausal complements. An example of *lician* in such a construction is given in (11) below:

(11) þa licade hire [...] þæt heo wolde þa baan up adon

then pleased(3-SG) her(DAT-SG) that she wanted the bones up take

‘then it pleased her to take up the bones’ (Alfred, *Bede* 4, 10)

While the construction of *lician* in (10) is generally considered personal due to the presence of a nominative constituent, the absence of case marking on the clausal constituent in (11) and the absence of a formal subject *hit*, which can be motivated by the presence of an experiencer argument in conjunction with a temporal adjunct in initial position, permit the analysis of the latter construction as impersonal according to the requirements discussed so far. The main factor that ties these instances together is the presence of an oblique experiencer argument, which is defined, in the present context, as an experiencer argument bearing any form of non-nominative case. Such a constituent is equally present in the case of one-argument experiencer verbs like *þyrstan* and *hyngrian* and at least optionally present in the case of verbs of happenstance like *gelimpan*, even though the latter appear to require a wider definition of the role of experiencer in order to qualify for this definition. The oblique case marking of the experiencer argument presumably indicates its non-agentive involvement in situations that express emotional and physiological sensations or chance events, whose exact cause typically remains equally obscure. The conceptual difficulty of identifying the cause of emotions or chance events is comparable to the difficulty of identifying an agent or force behind the natural phenomena discussed above, although these differ from the aforementioned situations in typically lacking an experiencer argument. It nevertheless seems that the definition of impersonal constructions as subjectless and their definition based on the presence of an oblique experiencer argument reflect similar conceptual premises, even if they extend to different sets among the verbs under discussion. Before comparing the scope of the two definitions, the potential subject status of the experiencer argument itself will be discussed, which constitutes the final theoretical option of a subject constituent in impersonal constructions.

2.1.1.1.4. Oblique Experiencers

2.1.1.1.4.1. Position

The experiencer argument of verbs like *þyrstan*, *hyngrian* and *lician* has itself been considered the subject of impersonal constructions, and the interpretation of these constructions as subjectless thus remains debatable even when there is no additional nominative complement present. The most adamant proponent of this view is Allen (cf. e.g. 1995, 1986a), but also Elmer

(1981: 48) remarks on the “double nature” of the experiencer argument, which he calls a “pseudo-subject” in the context of verbs with clausal complements (cf. also Ogura 1990: 33, who calls it a “‘subject’ equivalent”). For Elmer, the designation as “pseudo-subject” captures the fact that the experiencer in impersonal constructions with a clausal complement is typically preverbal and thus interpretable as subject based on the assumption of unmarked SVO word order in OE declarative main clauses, while, in semantic terms, it constitutes the (indirect) object. Elmer bases the latter interpretation of the experiencer on its alternative realisation by a PP in expressions like *it seems to me*, although oblique case marking by dative or, less frequently, by accusative case seems to point in a similar direction. If the preverbal position of the experiencer is taken as an argument in favour of its subject analysis, on the other hand, it should also be taken into account that instances with a clausal complement, which does not normally occur in preverbal position on account of its relative heaviness, may present a special case. Elmer (1981: 68) actually adopts an object analysis for experiencers in constructions with a postverbal nominative complement. The initial position of the experiencer in such cases is explained as topicalisation resulting from what Elmer calls the “animateness target” of Old English, by which he refers to a perceived tendency of animate constituents to precede other constituents. In terms of the semantic property of animacy, experiencers thus align with prototypical subjects.

2.1.1.1.4.2. Coreferential Deletion

There are further behavioural properties shared by oblique experiencers and nominative subjects in Old English, the most frequently discussed being coreferential deletion in coordinated main clauses (cf. e.g. Allen 1986a: 390, von Seeffranz-Montag 1984: 530f., Anderson 1984: 249). An example of control of deletion is presented in (12), which is taken from Old English, potential victimhood is exemplified in (13) and (14), which are both taken from Middle English:

- (12) ac **gode** ne licode na heora geleafleasht [...] ac asende him to fir of heofonum
 but God(DAT-SG)_i not pleased(3-SG) not their faithlessness but _____i sent(3-SG)
 them(DAT-PL) to fire from the-heavens
 ‘but their faithlessness did not please God and he sent them fire from heaven’
 (Ælfric, *Hom.* P. 20, 71)

(13) I wate þat þu has fastid lang and hungris nu

I know that you_i have fasted long and ____i hungers(3-SG) now

‘I know that you have fasted long and are hungry now’ (*Curs. M.* 12943)

(14) lewed men leued hym well and lyked his wordes

ignorant men_i loved him well and ____i liked his words

‘ignorant men loved him well and liked his words’ (Langland, *Piers Plow.*, *Prol.* 72)

In (12), the oblique experiencer argument *gode* ‘God’ is the coreferential antecedent of the deleted subject of the coordinated main clause, while in (13) and (14), the experiencer itself is deleted under coreference with the preceding subject. Note, however, that the *MED* (s.v. *hungren*) also contains a variant reading of (13), in which the verbal form *hongrest* is clearly marked for second person singular. Oblique case marking of deleted constituents is, of course, difficult to ascertain directly, and it is also not easily inferred for arguments of verbs that potentially occur in variable constructions. Given the diachronic change that impersonal verbs undergo during the ME period, this is a possibility with the ME examples, and no clear examples of victimhood to deletion seem to occur in Old English. The principal availability of oblique constituents for coreferential deletion is observable in modern languages such as Icelandic, where the regular case marking of deleted constituents can be established more readily on the basis of productive evidence. A constructed example of such a case is presented in (15) below:

(15) hún var mikill dýravinur og líkuðu hestar vel

she_i was a-big animal-lover and ____i pleased(3-PL) horses well

‘she was a big animal lover and horses strongly appealed to her’

Based on external evidence, the deleted constituent in (15) can be recovered as pronominal *henni* ‘her’ in the dative singular. The deletion of this constituent illustrates that behavioural subject properties can apply to Icelandic constituents even when they are not marked by nominative case and do not control verb agreement. The latter is apparent from the fact that the verbal form *líkuðu* ‘pleased’ is plural and thus agrees with *hestar* ‘horses’ rather than with *henni*. While the establishment of verb agreement in the ME second example in (14) above poses difficulties of its own due to the ambiguity of the preterit form, the OE example in (12) presents comparable evidence of control of coreferential deletion by an oblique constituent.

An alternative assumption would be that pronominal subjects in Old English were generally open to deletion, regardless of their coordination with a coreferential subject. In order to test this possibility, Allen (1986a: 390–393) investigated the evidence of coordination in *Ælfric’s Homilies*. Her conclusion is that, while some instances of free deletion outside of coordinated clauses do occur, deletion is noticeably favoured under coordination with a preceding coreferential subject. Conversely, instances of deletion under coreference with a preceding object, although occasionally present, are exceptionally rare. For impersonal verbs in particular, a similar investigation conducted by Allen on the basis of Healey/Venezky’s *Concordance to Old English* indicates that the syntactic behaviour of oblique experiencers is more similar to that of subjects than that of objects with regard to deletion. Unfortunately, the number of occurrences that match the restrictions of Allen’s investigation is rather small, but the fact that 8 out of 11 instances control deletion while only 3 do not appears to present a clear tendency at least (cf. Allen 1986a: 393f.). Control of coreferential subject deletion can thus be regarded as an established behavioural property that oblique experiencers share with nominative subjects.

2.1.1.1.4.3. Raising

Another property that oblique experiencers and nominative subjects potentially share is their availability for raising in examples like (16) and (17):

(16) þa ongan **hine** eft langian on his cypbe

then began(3-SG) him(ACC-SG) again to-cause-longing(INF) for his kin

‘then he began to long for his native country again’ (*Blickl. Homl.* 113, 14)

(17) that made **me** to mete

‘that made me dream’ (Chaucer, *Poems Parl. of F.* 108)

Such examples are difficult to interpret, however. Anderson (1984: 249) argues that subject-to-subject raising cannot be safely established for (16) due to the retained accusative case marking of the experiencer, and potential subject-to-object raising in (17) is one of the contexts that van der Gaaf (1904: 32f.) cites as examples of ambiguous constructions in Middle English, which permit both a personal and an impersonal interpretation of the verb. The examples will not be considered any further here, since the precise nature of the assumed transformational derivation is not of primary concern for the definition of impersonal verbs. It should be noted, however,

that raising constructions at least potentially provide additional evidence for the subject status of oblique experiencer arguments.

2.1.1.1.4.4. Verb Agreement

Examples like (6) from Old English, but also (15) from Modern Icelandic, demonstrate that control of verb agreement is typically exercised by the nominative constituent, if present, and not by the oblique experiencer, although there are occasional exceptions. Finding clear examples of contrastive verb agreement is, of course, not an easy task in narrative texts written largely in the third person, since diagnostic cases necessarily involve a difference in grammatical number between the experiencer argument and the second argument. A potential instance is (18):

(18) **him** gelicade hire þeawas

him(DAT-SG) pleased(3-SG) her manners(NOM/ACC-PL)

‘her manners pleased him’ (*Chron.* Pl. 201)

Here, the second argument *þeawas* ‘manners’ is, in fact, ambiguously marked by nominative-accusative case, which is a general feature of the plural in all major OE declensions. Taking the form as representing nominative case, Allen (1986a: 388) concludes that nominative constituents could fail to trigger verb agreement in Old English, arguably because they did not constitute the subject in constructions like the one in (18) above (for the alternative interpretation as accusative cf. Fischer/van der Leek 1983: 347f., Anderson 1988: 17). Ogura (1990: 36), however, points to Mitchell’s discussion (1985: 10) of potential preterit plural forms with early loss of final *-n* already during the OE period. Although Mitchell finds alternative explanations for potential examples of this phenomenon in *Beowulf*, his overall conclusion is that the preterit ending *-e* is no unambiguous marker of the singular, which would render forms such as *gelicade* ‘pleased’ in (18) inconclusive. Similar cases of phonological weakening affect present-tense forms and ultimately lead to a merger of 3. pl. *-ap* > *-ep* > *-p* and 3. sg. *-ep* / *-p*, cf. Mitchell 1985: 636f. What is more, an occasional absence of verb agreement can also be observed for clear cases of postverbal subjects (cf. Mitchell 1985: 637, who attempts to explain at least some of these instances as collective nouns). Taken together, these observations suggest that the evidence of verb agreement with regard to the subject status of oblique experiencers is not particularly strong, a fact admitted also by Allen (1986a: 395).

Later instances of non-standard verb agreement in Middle English can be observed in the examples in (19) and (20), which present two different cases:

(19) **me think** we shal be strong enough (More, *Wks.* 1577, 1254 B)

(20) how **likes þe** me nowþe

how pleases(3-SG) you(OBJ-SG) me(OBJ-SG) now

‘how do you like me now’ (*Will. of Pal.*, 1740)

In (19), the first-person experiencer *me*, despite its oblique case marking, appears to be in control of agreement of the verbal form *think*, which lacks the third-person singular marker required by the clausal complement *we shal be strong enough*. In (20), both arguments are realised by oblique pronouns, viz. *þe* ‘thee’ and *me*, but neither of these is compatible with the third-person singular form of the verb *likes*. Such instances seem to be restricted to later periods of English, however. The ones in (19) and (20) above are dated to the early 16th and the mid-14th century respectively, which induces scholars such as Lightfoot (1979: 232f.) to maintain the possibility of a subject analysis of oblique constituents in the ME period. While positive examples of control of verb agreement exercised by the oblique experiencers remain sporadic and appear to be explicitly avoided by the conversion of expressions such as *me think* in (19) into instances of NPs *my think(s)*, *my thought(s)* (cf. Visser 1963: 25), “double-object constructions” like the one in (20) inevitably require a subject interpretation of either one of the constituents according to Lightfoot (1979: 234), who assumes a general subject requirement for canonical clauses. Other scholars usually take such double-object constructions as evidence of the transitory state of late ME (cf. e.g. Jespersen 1984: 221, van der Gaaf 1904: 68, also Elmer 1981: 152f.) and regard them as blends of personal and impersonal constructions. In a slightly different vein, Allen (1986a: 398) argues that the assignment of dative case to postverbal constituents proceeded by means of structural assignment at this stage, while subject pronouns could still receive dative case through lexical assignment, and the instances of double-object constructions are thus interpreted as the grammatically regular expression of the simultaneous manifestation of both processes. However, the overwhelming evidence of verb agreement being controlled by nominative constituents throughout the historical stages of English makes it difficult to regard individual examples such as (19) and (20) as absolute proof that control of verb agreement had been acquired by oblique experiencers as a regular subject property, which remains a rather unlikely possibility at best.

2.1.1.1.4.5. Case Marking

The strongest evidence against the subject status of oblique experiencer arguments is usually perceived in their oblique case marking, since prototypical subjects of Old English are typically marked by nominative case (cf. Mitchell/Robinson 2007: 105, for an application to impersonal verbs cf. e.g. Fischer/van der Leek 1987: 82f., Anderson 1988: 9). The preceding discussion has illustrated, however, that oblique constituents can exhibit behavioural subject properties such as preverbal position and control of coreferential deletion in Old English as well as Icelandic, despite their lack of equivalent coding properties such as nominative case and control of verb agreement. Further evidence of behavioural subject properties is presented by Anderson (1988: 243–247) based on the discussion in Cole et al. (1978) of Kannada, a Dravidian language of southern India, and Hindi-Urdu, an Indo-Aryan language. In both of these languages, oblique constituents occupy the subject position and control reflexivisation and coreferential deletion in coordinated main clauses, but victimhood to deletion is restricted to main clauses in Kannada, while, in Hindi-Urdu, it is restricted to embedded clauses (“deletion under equi”). Apart from the problems this uneven distribution of behavioural properties poses to the distinction of subjects and non-subjects, it raises the general question of whether coding properties such as nominative case marking and control of verb agreement need to be viewed as necessary criteria of subjecthood.

Allen (1986a: 381) argues against this view, pointing to the apparent existence of oblique subjects in a range of other languages, including Japanese, Korean and Georgian (cf. Allen 1986a: 389). The facts from these languages are intended to illustrate the dissociation of syntactic subjecthood and case marking. In Georgian, for example, oblique arguments of so-called “indirect verbs” generally control verb agreement, a feature which is comparable to the sporadic control of verb agreement observed of oblique experiencers in Middle English (cf. also McCawley 1976: 201f.). The long-term diachronic development of English impersonal verbs could thus be conceptualised as a gradual acquisition of subject properties by the experiencer argument, starting with its preverbal position, which was favoured by the experiencer’s typical animacy and a concomitant tendency towards topicalisation, continuing with behavioural properties like control and, potentially, victimhood to coreferential deletion, and, in later stages, sporadic instances of verb agreement, and ultimately resulting in the acquisition of nominative case marking, which is the most typical case for experiencer arguments in Present-Day English. The definition of impersonal constructions as subjectless would then depend partly on the question of where to draw the line between subject-like properties and subjecthood proper.

2.1.1.2. Oblique-Experiencer Verbs

While the criteria of the standard definition that was laid out at the beginning of this section and the related classification of impersonal constructions as subjectless may have seemed straightforward initially, the theoretical definition of subjects is arguably more complex, and, depending on its exact formulation, can lead to the disqualification of the majority of patterns that verbs commonly discussed in the context of impersonal constructions normally enter into. The presence of an oblique experiencer argument, on the other hand, presents a criterion that subsumes the majority of constructions, regardless of the presence of a second argument and its realisation by clausal or nominative constituents, and regardless of the presence or absence of an expletive. A comparison of the scope of these two definitional approaches with regard to the types of verbs that have been considered so far leads to the following observations.

First, the seemingly prototypical weather verbs, for the most part, escape either definition, at least if only the historical stages of English are concerned, since weather verbs do not normally take an experiencer argument and are most frequently attested with an expletive pronoun already in Old English. The reason that weather verbs have been considered “true” or “really impersonal” verbs thus appears to lie in their semantic argument structure, which does not profile any regular participants on account of the inherent difficulty of identifying an agent or force behind the natural phenomena that are expressed, while the syntactic realisation of this frame typically requires the provision of a semantically empty constituent. Second, one-argument experiencer verbs such as *þyrstan* and *hyngrian* can generally be subsumed under either definition, since they more frequently occur without an expletive, and since they profile an experiencer argument that is marked by oblique case in earlier stages of English. The oblique case marking of the experiencer argument appears to reflect a similar conceptual premise compared to weather verbs, since the precise causes of the physiological sensations that affect the experiencer are not easily identified. Third, verbs of happenstance like *gelimpan* are not always subjectless in their one-argument use, which frequently sees the addition of an expletive pronoun, but they mostly qualify as subjectless when they occur with an additional experiencer argument, since this appears to restrict the occurrence of the expletive more fully. The typical realisation of the second argument by a clausal constituent is compatible with the notion of subjectless constructions, since clausal constituents are generally considered to be ambiguous in their subject status. In terms of the definition as oblique-experiencer verbs, verbs of happenstance are certainly not a prototypical case, and they only fall into this category in as much as an oblique experiencer argument is optionally present, although it is less frequent than with other experiencer verbs. The reason for this optionality is that chance events can be

expressed without regard for their effect on an experiencer, a term that applies to such participants only in a wider sense, while the experiencer of physical sensations or emotions is logically presupposed by the verbal semantics and compatible with a narrower definition of the term. Verbs of emotion such as *lician* constitute prototypical examples of verbs that potentially occur with an oblique experiencer argument in earlier stages of English, but the classification of their construction as subjectless largely depends on the given complementation pattern. Instances with a nominative complement are generally excluded from the definition, and instances involving clausal complements depend on the presence or absence of an expletive for their analysis.

The syntactic difference between constructions with and without an expletive that can be observed for weather verbs and the use of verbs of happenstance with a single argument as well as a range of other verbs involving clausal complements is an interesting phenomenon in its own right, and the diachronic distribution of expletive pronouns and the implication of their distribution for the grammatical system of English are definitely worthy of research. When it comes to impersonal constructions, however, it seems unfortunate to restrict their definition to instances without an expletive or an additional nominative constituent. First of all, this distinction does not capture the cognitive basis that appears to be underlying all of the verbs discussed so far. The most obvious case are weather verbs, which entirely lack a participant that could be coded as subject, but a similar point applies to the participants of verbs denoting physical sensations or emotions, which are semantically less suitable as subjects. In either case, the presence of an expletive merely constitutes an alternative grammatical expression of this conceptual premise and does not affect the semantic frame. For verbs with clausal complements, the occurrence of an expletive also appears to be at least partly restricted by the presence of other constituents in that typical experiencer verbs and instances of verbs of happenstance that include an additional experiencer argument are much less frequently accompanied by expletives. This seems to be attributable to the fact that the presence of two verbal arguments alongside optional adjuncts leaves less room for syntactically vacant slots that could potentially require an expletive, and this, in turn, seems to imply that, while instances of apparent vacancies in the syntactic structure can be regarded as a positive indication of impersonal constructions according to the definition as subjectless, instances without such vacancies are, in fact, not diagnostic. It appears that the morphological ambiguity of clausal constituents is the most conducive to the analysis of such instances as subjectless, but this does not seem justifiable with reference to distributional properties, since the restriction of clausal constituents in this regard can be motivated by their relative heaviness. If this fact is taken into

account, the theoretical distinction between instances with clausal complements as impersonal and parallel instances with nominal complements marked by nominative case as personal seems ill-founded.

While the syntactic requirement of expletive constituents in the absence of semantic arguments appears to be largely in place already in the OE period, the replacement of oblique experiencer argument by nominative experiencer argument is a process which, for the most part, unfolds during the historical stages of English. The two different approaches discussed in this section of defining the distinctive features of impersonal constructions are thus not mutually exclusive, but an emphasis on the oblique case marking of the experiencer argument is simply more relevant for the purposes of the present study. The subsequent sections 2.1.2 and 2.1.3 will provide a more systematic overview of the different semantic and syntactic classifications that have been applied to impersonal verbs in general and to two-argument experiencer verbs in particular, which will be followed in section 2.1.4 by a discussion of some of the more important theoretical treatments of impersonal verbs in terms of synchronic grammar.

2.1.2. Semantic Classification

2.1.2.1. Wahlén 1925

Given the apparent difficulty of defining the scope of impersonal constructions, it is not surprising that multiple versions of their synchronic classification in terms of lexical semantics and syntactic complementation have been put forward as well (for an overview of the most frequently discussed features cf. Traugott 1992: 208–213). Since many theoretical treatments tend to adopt a more eclectic focus, it is well worth revisiting earlier treatments by scholars such as Wahlén, whose meticulous account remains undaunted by the apparent diversity of eligible verbs and constructions. On the broadest level, Wahlén (1925: 11–12, followed by Mitchell 1985: 428), makes a tripartite distinction between impersonal constructions in the active voice, impersonal constructions in the passive voice, and impersonal constructions containing the copular verbs *beon*, *wesan* and *weorþan*, cf. e.g. *laþ beon* ‘to be odious, hateful’ or *leof beon* ‘to be dear, agreeable’. The two latter types of constructions are beyond the focus of the present investigation as well (for a discussion of impersonal and recipient passives cf. Allen 1995, for a discussion of “phrasal impersonals” cf. Denison 1990: 125, the general account in van der Gaaf 1904, and the separate treatment in van der Wurff 1992). Regarding impersonal constructions in the active voice, Wahlén distinguishes eight semantic classes,

which comprise 1. expressions for natural phenomena, 2. expressions for specifications of time, 3. expressions denoting locality, 4. expressions of physical and mental affections, 5. expressions of mental activity, 6. expressions denoting the course of events, 7. expressions implying a statement or explanation, and 8. the *mæg-cunnian*-group. While some of these will be familiar from the previous section, others may require further exemplification and also allow for some interesting points to be made.

The semantic classes 1, 2 and 6 roughly correspond to the previously illustrated types that were represented above by *sniwan* ‘to snow’, *nighten* ‘to grow dark’ and *gelimpan* ‘to happen’. For the last of these, the term “happenstance” has been used instead of the term “course of events”. Of special interest is Wahlén’s distinction between “physical and mental affections” (class 4) on the one hand and “mental activities” (class 5) on the other, despite some semantic and constructional overlap (Wahlén 1925: 38, 58f.). The physical affections of the former class include *þyrstan* ‘to thirst’ and *hyngrian* ‘to hunger’, but also *calan* ‘to grow cold’, while mental affections are represented by *lician* ‘to please’, *hreowan* ‘to rue’ and *sceamian* ‘to feel ashamed’. Within the latter subgroup, Wahlén makes an even more fine-grained distinction between “personal” and “sympathetic mental affections”, the latter of which comprise a rather small group represented by *ofhreowan* ‘to cause pity’, whose primary semantic characteristic is the typical involvement of a second animate participant. The “mental activities” of class 5, on the other hand, include *mætan* ‘to dream’, *þyncan* ‘to seem’ and *tweogan* ‘to cause doubt’. Examples of these verbs are given in (21)–(23) below:

(21) hit gelamp þæt **hine** mætte

it happened that him(ACC-SG) dreamt(3-SG)

‘it happened that he had a dream’ (Ælfric, *Gen.* 37, 5)

(22) **me** nu þynceþ & bet licaþ, þæt [...]

me(DAT/ACC-SG) now seems(3-SG) and better pleases that

‘it seems to me now and pleases me more that’ (Alfred, *Bede* 1, 27)

(23) **nænne mon** þæs tweogan ne þearf

no-one man(ACC-SG) that(GEN-SG) not cause-doubt(INF) may(3-SG)

‘no man may doubt that’ (Alfred, *Boeth.* 26, 12)

The semantic difference between such “mental activities” and the “mental affections” of class 4 may seem insignificant at first, but the term used for the former appears to indicate that they involve a more active type of mental process. Pocheptsov (1997: 477) considers the idea that mental affections do not involve “an assessment of facts in terms of ‘pleasant/unpleasant’” and that they can thus be conceived of as more neutral events. Yet evaluative meanings are easily derived from the basic meanings by the addition of a prefix to the simple verbs, an example of which is *offyncan* ‘to cause regret, displeasure, offence’, and they can also be inherent in the simple verbs as demonstrated by the coordination of *byncep & bet licap* ‘seems and pleases better’ in (22) above. Pocheptsov’s own distinction of “physical feelings”, “emotional feelings and states” and “physical and mental actions and states” (Pocheptsov’s classes 1–3, cf. 1997: 471–477) does not help much to clarify the situation, since it contains a similar overlap of different semantic concepts. The crucial point, however, seems to be that “mental activities” comprise verbs of cognition, which arguably represent a more active process and are thus distinct from physiological sensations and emotions.

Wahlén’s remaining classes 3, 7 and 8 are relatively infrequent. Class 3 contains only three verbs *cuman*, *becuman* and *gegan*, which, on their own or in combination with a locational adjunct, all mean something like ‘to go to, arrive at’. They are commonly used to translate the impersonal passive of Lat. *pervenitur* ‘it was reached’ or similar expressions (cf. Wahlén 1925: 36f.), in which case they are rather comparable to the indefinite type of reference discussed above, which is represented by indefinite pronouns such as OE *man* ‘one’. Note also that these verbs have a figurative sense ‘to come to pass’, which belongs more properly into class 6 of verbs of happenstance. Expressions implying a statement or explanation (group 7) can be compared to impersonal and recipient passives, cf. the example in (24):

(24) on þisum bocum **us** segþ þæt [...]

in these books us(ACC/DAT-PL) says(3-SG) that

‘it tells us in these books that’ (Ælfric, *De Vet. Test.* 6, 41)

This particular construction is a reminder of the exegetic nature that is common of most early English writings, although it is quite rare according to Visser (1963: 7, 39f.). In the active voice, expressions like the one in (24) can be regarded as semantically similar to existential clauses, in which the verb *secgan* ‘to say’ or, similarly, *cwēpan* ‘id.’ and *cyþan* ‘to proclaim’ fulfils the function of an existential verb. The difference, however, is that a recipient or addressee such as *us* may be present, although examples without are more numerous in Visser’s list (cf. Visser

1963: 7, 39f). The expression also involves a second complement, which can be analysed as theme in terms of its semantic role. The suppression of an agent, which is the typical function of passive constructions, thus appears to be a grammatical phenomenon that is not systematically linked to the semantics of verbs of speaking. Finally, class 8 does not actually constitute a semantic unit according to Wahlén (1925: 105), but is tied together by the occurrence of its members as verbal complements of other verbs in impersonal constructions, including the eponymous *mæg cunnian* ‘it may test, examine’. Each of the four constructions cited by Wahlén only has a single occurrence in his corpus, and the combination of the relevant verbs with *magan* ‘to be able, may’ and, once, *sculan* ‘shall’ seems to render them expressions of happenstance, which can be associated with Wahlén’s class 6.

2.1.2.2. Elmer 1981

A more concise and more frequently cited categorisation of impersonal verbs was developed by Elmer (1981: 6), who distinguishes five semantic classes: 1. *rue*-verbs, 2. *please/desire*-verbs, 3. *behave*-verbs, 4. *happen*-verbs, and 5. *seem*-verbs. An important factor in this more economical treatment is his focus on verbs with two arguments, which disregards weather verbs and expressions of time (Wahlén’s classes 1 and 2), the marginal cases (Wahlén’s classes 3, 7 and 8), but also “physical affections” like *þyrstan* and *hyngrían* from Wahlén’s class 4, which typically occur with a single experiencer argument only. The “mental affections” (or “emotional states” in Pocheptsov’s terminology) are divided into *rue*-verbs on the one hand and *please/desire*-verbs on the other. This could be thought of as a fundamental distinction between negative and positive emotions, although differences in the respective complementation patterns of these verbs, which will be discussed further below, also appear to play a role for Elmer’s distinction. In addition, Wahlén’s group 6 indicating the “course of events” is, quite suitably, divided into verbs of obligation (*behave*-verbs) and verbs of happenstance (*happen*-verbs), as these present clear semantic differences. The restriction of Elmer’s classification to impersonal verbs with two arguments is thus counterbalanced by a more fine-grained semantic distinction of the remaining cases. The only representative of the final class 5 of *seem*-verbs that is discussed by Elmer is OE *þyncan*, as ME *seem* itself is only derived later in English from ON *sæma* ‘befit’, itself a *behave*-verb. It is unfortunate that other verbs from Wahlén’s class 5 of “mental activities” like *mætan* and *tweogan* are not considered in Elmer’s treatment, as a fuller comparison of verbs of cognition and verbs of emotion would have been welcome. It should also be noted that additional verbs of cognition like ME *remembren* ‘to remember’ (from

OF *remembrer* ‘id.’) enter impersonal constructions during later stages of English (cf. McCawley 1976: 194). The same is true of verbs referring to (negative) possession or necessity like ME *lakken* ‘to be lacking’ and ME *wanten* ‘to be in need, lack’ (from ON *vanta* ‘to lack’), which constitute an additional semantic class. Elmer’s semantic classification of impersonal verbs thus cannot be expected to cover the full semantic range of verbs in impersonal constructions due to its limitation to Old English (for other languages like Russian and Japanese, dynamic modality appears to be relevant to impersonal constructions as well, cf. McCawley 1976: 194, similarly von Stefranz-Montag 1984: 524; for semantic classes in EME, cf. Rissanen 1999: 250).

2.1.2.3. Möhlig-Falke 2012

The common restriction of the discussion of impersonal verbs to experiencer verbs with two arguments, which applies not only to Elmer’s treatment but also to e.g. Fischer/van der Leek 1983, Allen 1995, Loureiro-Porto 2009, Miura 2015 and the present study), is also noted by Möhlig-Falke (2012: 21), who critically remarks that the frequent focus on “mental verbs” has led to a theoretical overgeneralisation of the role of experiencer. In her own treatment, she distinguishes eight lexical fields, viz. 1. physical sensation, 2. emotion, 3. cognition, 4. existential experience, 5. motion, 6. appropriateness, 7. (non-)availability, and 8. benefaction. These are arranged into a taxonomy of conceptual domains that comprises physical sensation, mental experience (including emotion, cognition and perception) and existential experience under the more general domain of human experience besides the conceptual domains of motion and possession (cf. Möhlig-Falke 2012: 52–56). Although Wahlén is not cited in this context, the domain of motion appears to be equivalent to his class 3 of verbs denoting locality. This class is not attributed primary importance by Möhlig-Falke, however, which means that the domains of human experience and possession remain the most relevant to the discussion of impersonal verbs. With regard to their frequency, she notes that verbs of physical sensation and emotion exhibit a larger type frequency, while verbs relating to cognition, ownership and appropriateness exhibit a much larger token frequency in her corpus of Old English than would be expected on the basis of their limited type frequency. This finding appears to be confirmed by Miura’s discovery of several isolated instances of impersonal constructions in most of the semantic subgroups of verbs of emotion that she distinguishes (cf. Miura 2015: 56–94). The additional recognition of verbs in the domain of possession compared against Wahlén’s and Elmer’s classifications appears to be a result of the fact that the latter studies are based on Old

English, while verbs relating to necessity or (non-)availability such as ME *lakken* and ME *wanten* are commonly found in the ME period only. Its conceptual separation from core experiencer verbs relating to human experience seems justified, even though a situation of lack or want might also be conceived of as an experience when it is related to a sentient entity that is affected by it. The subsumption of verbs of happenstance under the concept of human experience, on the other hand, while providing a convenient form of terminological consistency, appears to disregard the optional involvement of the experiencer in the former type of situation, which was noted above in connection with the frequent instances of *gelimpan* that do not involve an experiencer argument.

2.1.2.4. Polysemy and Semantic Changes

As a final point, it should be noted that impersonal verbs may comprise multiple senses and thus belong to different semantic classes depending on which sense is applicable at a given time. The eponymous verb of Elmer's *behove*-class, for example, expresses non-availability in OE *behofian* 'to need', a sense retained in Swedish *behöva* 'id.', itself a loan from Middle Low German *behoven*, and the verb only acquires the sense of moral obligation in English during the 12th century (cf. Elmer 1981: 40f.). This development can be conceived of as a shift from "internal necessity" towards "external necessity" (cf. Allen 1997: 10, similarly Loureiro-Porto 2009: 109), which arises from the basic sense by implication of social or other external standards. Another *behove*-verb, *gebyrian* 'to pertain to, befit', can also mean 'to happen', in which case it belongs to the *happen*-class (cf. Elmer 1981: 40¹). The latter sense is probably basic (cf. ON *byrja* 'to begin'), although the former is also represented by OHG *giburren* 'to happen, pertain to'. A transitional meaning would be 'to (rightly) happen to', which seems to be marginally attested for *gelimpan* 'to happen' as well, which glosses Lat. *quadrāre* 'to be suitable'. Finally, ME *semen* 'to seem, to be fitting' (cf. ON *sæma* 'to befit') occurs in alternative senses that align it either with the *seem*-verb *þyncan* 'to seem' or with the *behove*-class. A transition from a more general meaning 'to seem' to a more specific meaning 'to seem good/appropriate' is conceivable and attested in Lat. *vidētur*, which covers exactly these meanings. Yet in Old Norse, the more specific sense of the verb appears to be primary, cf. *sæmd* 'honour' and the related verb *sæma* 'to honour, to reconcile', also in OE *seman* 'to reconcile, settle'. The development of ME *semen*, which continues the deontic sense of ON *sæma* rather than the more concrete sense of OE *seman*, must then be viewed as a process of generalisation 'it seems good' > 'it seems' (thus van der Gaaf 1904: 22f.).

2.1.3. Syntactic Classification

2.1.3.1. Ogura 1986

A discussion of the syntactic patterns that impersonal verbs enter requires a basic distinction between impersonal verbs and impersonal constructions, which was already implicit in the preceding sections, since some of the observable patterns are not compatible with either of the two definitions of impersonal constructions that were discussed above. The designation of a verb as impersonal thus needs to be understood as referring to its potential of being used in at least one of the constructional patterns defined as impersonal, which can, of course, only be judged by its extant attestations. The most comprehensive classification of such patterns is probably the one by Ogura (1986: 38f.), which appears to be based on Mitchell (1985: 428). It includes impersonal passives as well as phrasal impersonals, but, more importantly, it also makes a distinction within impersonal constructions in the active voice between zero-argument verbs on the one hand and one- and two-argument verbs including an oblique experiencer on the other (the oblique experiencer is referred to as a “dative or accusative of person” in Ogura’s terminology). The presence of “formal *hit*” is permitted in either case. In her more recent monograph on expressions of emotion, Ogura (2013: 101f.) distinguishes several syntactic environments for OE impersonal and related constructions, which include predicates without nominal complements and an optional clausal complement, predicates with an experiencer, with *hit*, with an experiencer and *hit*, with provisional *þæt* and *þis* introducing a clausal complement, with either of these accompanied by an additional experiencer, and, finally, with a nominative of thing (i.e. an inanimate nominal complement), either with or without an experiencer. The large number of distinctions demonstrates the often less than clear-cut diversity encountered in natural languages. It is less functional, however, in providing a systematic overview that can be used as the basis of further research. Despite the existing diversity, a more economical treatment thus seems preferable and is, indeed, available from Elmer’s study (1981), which obviously benefits in terms of clarity from the fact that it focusses on two-argument verbs.

2.1.3.2. Elmer 1981

2.1.3.2.1. Type S

Elmer’s starting point in his syntactic classification of impersonal constructions is the construction with sentential complements, a designation that he uses as a cover term for both finite and non-finite complement clauses, although the former clearly dominate his discussion.

Elmer claims that this construction is “highly representative in terms of occurrence” and “shows significant syntactic patterning” (1981: 21), by which he presumably refers to the fact that clausal complements occur with virtually all of the verbs under discussion (cf. his overview 1981: 46, in which the only exception is presented by intransitive *lustfullian* ‘to rejoice’), while the occurrence of nominal complements is restricted to *rue*- and *desire*-verbs (but not *please*-verbs), at least for nominal complements in the genitive case. Nominal complements occur more widely if instances in the nominative are included, but the relevant constructions are no longer subjectless and thus not strictly speaking impersonal according to the definition most commonly adhered to. This problem is gracefully avoided by using constructions with a morphologically ambiguous clausal complement as the basic model. The issue of constituent order is also more straightforward in these instances, since finite clausal complements consistently and non-finite clausal complements most frequently occur in clause-final position, leaving the preverbal position to the experiencer argument. Overall, there are thus a number of specific restrictions to what Elmer considers the “basic structure” of impersonal constructions (cf. Elmer 1981: 14, 23) which do not apply to all of its variations. Nevertheless, it can serve as a convenient starting point for the description of the observed types.

Specifically, Elmer distinguishes three main variants of the type with clausal complements, which he refers to as type S, the personal type, and the *it*-construction. These variants are exemplified by instances of *rue*-verbs in (25) through (27), adapted from Elmer’s (47), (49) and (54):

- (25) **me ofhreow** þæt hi ne cuþon þa godspellican lare
 me(ACC/DAT-SG) rued(3-SG) that they not knew the Evangelical lore
 ‘it grieved me that they did not know the Evangelical lore’ (Ælfric, *Hom.* Th. 1, 2, 22)
- (26) gif **we** þonne **scomiaþ** þæt we [...]
 if we(NOM-PL) then are-ashamed(1-PL) that we
 ‘if we are ashamed then that we’ (Alfred, *C. P.* 63, 5)
- (27) þa **ofþuhte** þæt **Mariuse** [...] þæt [...]
 then offended(3-SG) that Marius(DAT-SG) that
 ‘then it offended Marius that’ (Alfred, *Oros.* 236, 3)

Example (25) presents the basic type S with a preverbal oblique experiencer *me* and a postverbal complement clause as the second argument. In (26), the experiencer argument, by contrast, is realised by the nominative pronoun *we*, and the verbal form *scomiāþ* shows person and number agreement. This type of personal construction is not common in Elmer’s collection of OE examples and not established for any other *rue*-verb apart from *sceamian*, but it entails a significant complication for the analysis, since *sceamian* is also attested in impersonal constructions in Old English. The distribution of the impersonal type S and its personal variant is thus not lexically restricted, and it is also not perfectly separated in diachronic terms, since personal and impersonal constructions are attested alongside each other already in Old English. It should also be noted that personal constructions of *rue*-verbs occur more widely in Old English outside of constructions with clausal complements, which include constructions with a second nominal complement in the accusative like the one in (28):

(28) **hie** ne magon ealneġ ealla on ane tid emnsare hreowan

they(NOM-PL) not can(3-PL) always all(ACC-PL) at one time equally-bitterly repent(INF)

‘they cannot at once repent of all [sc. sins] equally bitterly’ (Alfred, *C. P.* 413, 28)

In this example, the combination of a nominative constituent *hie* ‘they’ and an accusative constituent *ealla* ‘all’ formally results in a transitive construction of the verb *hreowan*. While Elmer’s type S with oblique experiencers appears to be the most frequent type among constructions with a clausal complement, this definitional restriction plays a role in limiting the observed variation with nominative experiencer arguments. Instances of the third variant with expletive *hit* are also virtually absent from Elmer’s sample of OE *rue*-verbs, although a number of examples involving “complex predicates” (i.e. phrasal impersonals) like *sorhlic beon* ‘to be grievous’ do exist. The example of this construction supplied in (27) actually contains the demonstrative pronoun *þæt*, which can be regarded as more referential than the anaphoric pronoun *hit* (on the notion of gradient referentiality cf. Denison 1993: 97, 1990: 132). The use of the demonstrative is distinguished as a separate type by Ogura (cf. above) and also by Allen (1995: 87f.), who refers to it as “type DEM”. Allen makes similar observations about the infrequency of *hit* with clausal complements prior to the 11th century, but she also points out that *hit* occurs frequently in constructions without an experiencer like *þa gelamp hit þæt [...]* in (8) above, which means that the *it*-construction, like the personal type, is rare in Old English only if it is defined in terms of constructions with two arguments.

2.1.3.2.2. Type N

For constructions with two nominal complements, Elmer makes a distinction between a basic type N and two main variants, types I and type II. These can be exemplified, once more, with the help of *rue*-verbs, which are cited below in (29)–(31), corresponding to Elmer’s (1), (32) and (40) in his chapter II:

(29) **him** ofhreow þæs mannes

him(DAT-SG) caused-pity(3-SG) the man(GEN-SG)

‘he felt pity for the man’ (Ælfric, *Hom.* Th. 1, 192, 16)

(30) þylæs [...] **him** þonne gehreowe sio ælmesse

the-less him(DAT-SG) then rue(3-SG-SUBJ) the almsgiving(NOM-SG)

‘lest he then regret his almsgiving’ (Alfred, *C. P.* 324, 7)

(31) þa **se mæssepreost** þæs mannes ofhreow

then the priest(NOM-SG) the man(GEN-SG) caused-pity(3-SG)

‘then the priest felt pity for the man’ (Ælfric, *Hom.* 2, 142, 262)

In contrast to type S, the impersonal nature of type N in (29) is explicitly manifested in the non-nominative case marking of the two nominal constituents, which renders the construction subjectless according to most analyses. The preverbal experiencer is marked by dative case and the postverbal argument by genitive case. Both types of arguments are also attested in the nominative, however, as illustrated by the postverbal nominative argument *sio ælmesse* ‘the almsgiving’ in (30) and the nominative experiencer argument *se mæssepreost* ‘the priest’ in (31). Elmer considers the latter variant to be structurally equivalent to the personal variant of type S. There are potential differences in terms of constituent order, however, since both the experiencer argument and the second argument occur preverbally in the example above. Regarding the former variant, Elmer (1981: 67f.) makes an additional distinction between type Ia and type Ib, depending on whether the second nominal constituent (type Ia) or the experiencer argument (type Ib) precedes the other in the constituent order. This formulation is deliberately phrased without reference to the position of the verb, which can be clause-final in Old English, in which case the two arguments are not distinguished by pre- or postverbal position. Additional examples of type I are given in (32) and (33), which correspond to Elmer’s (34) and (28) and illustrate type Ia and type Ib respectively:

(32) þes siges gewearþ **Punicum**

this victory(NOM-SG) happened(3-SG) the-Carthaginians(DAT-PL)

‘this victory occurred to the Carthaginians’ (Alfred, *Oros.* 176, 4)

(33) hu **him** se siges gelicade

how him(DAT-SG) the victory(NOM-SG) pleased(3-SG)

‘how the victory pleased him’ (Alfred, *Oros.* 156, 25)

In his discussion of word-order patterns, Elmer (1981: 68) analyses type Ia as SVO and type Ib as OVS, which presents little difficulty in case of the former, since it is the basic order in OE declarative main clauses. The alternative constituent order in type Ib is motivated by Elmer’s animacy target, by which he intends to capture the perceived tendency of animate constituents to occur towards the beginning of a clause. According to Elmer, this tendency results in “the appearance of a subjectless structure [...], although there is a syntactic subject in postverbal position” (1981: 68). This statement demonstrates, once again, that the conceptual boundaries between impersonal and personal constructions are not as clear-cut with nominal complements as they appear to be with clausal complements, which are typically restricted in their distribution to postverbal position. Finally, a third subtype *it-V-NP* is posited by Elmer (1981: 71) as a variation of type Ia with pronominal *hit* instead of a nominal constituent in initial position. Elmer’s discussion makes it clear, however, that *hit* presents a referential pronoun in this subtype and does not correspond to expletive (or “provisional”) *hit* in the *it*-construction with clausal complements. The point is illustrated by constructed examples only, but the instance in (34) appears to be a suitable example:

(34) hit licode **herode**

it pleased(3-SG) Herod(DAT-SG)

‘it pleased Herod’ (*OE Gosp. Mt.* 14, 6)

The fact that expletive *hit* does not seem to be attested alongside two nominal complements matches Allen’s observation regarding its infrequency in *it*-constructions involving an experiencer and a clausal complement. The occurrence of expletive *hit* thus appears to be largely excluded in the presence of two semantic arguments, regardless of whether the second argument is realised by a nominal or a clausal complement. A variant of type II with a nominative experiencer and expletive *hit* is, likewise, not recorded by Elmer.

2.1.4. Theoretical Accounts

While the classifications of impersonal constructions that were surveyed in the preceding paragraphs deserve credit for their faithful representation of the empirical evidence, they make few assumptions in terms of a more rigid theoretical formalisation of language, which is usually considered to be a prerequisite to any explanation of diachronic syntax (cf. e.g. Fischer 2007: 14–17, 1992: 236). An awareness of the existing variety of impersonal constructions is important, however, in order to be able to judge the scope of one’s theory and to avoid restricting oneself to a small set of streamlined data, which can be an inherent danger of more theory-driven work (cf. also a rather sceptical remark regarding the suitability of such treatments by Ogura (1990: 31)). Nevertheless, the work of scholars such as Lightfoot (1979 and later publications) as well as that of Fischer/van der Leek (especially 1983, also 1987) has been extremely influential and stimulated a great deal of discussion both with regard to the synchronic analysis and the diachronic treatment of impersonal verbs. The synchronic account by the latter two authors will be discussed in the remainder of this section, while that of the former will be taken up in section 2.2 on diachronic change.

2.1.4.1. Government and Binding

2.1.4.1.1. Single-Entry Approach

Fischer/van der Leek’s account is based in the general framework of Government and Binding and focusses on the lexical entries of impersonal verbs. The authors distinguish three basic constructional types, viz. type i, ii and iii (cf. 1983: 347–354), which closely resemble Elmer’s type N and its variants type I and type II. Note the difference in numbering, however, which may lead to terminological confusion, since their type ii corresponds to Elmer’s type I, and their type iii corresponds to Elmer’s type II. Fischer/van der Leek’s more explicit terms “impersonal construction”, “cause-subject construction”, and “experiencer-subject construction” may therefore be preferable. These assign the semantic role of cause to the relevant argument of two-argument verbs in a similarly generalising way that the term experiencer is commonly used for the other argument. Also, the potential analysis of oblique constituents as subjects on account of their behavioural properties, which Fischer/van der Leek apparently do not subscribe to, would suggest the use of less contentious designations for the latter types as “nominative-cause construction” and “nominative-experiencer construction”. Following Anderson (1986: 170f.), the three patterns are usually illustrated with examples of the single verb *ofhreowan* (cf.

e.g. Denison 1990: 112f.), two instances of which have already been cited as (29) and (31) above, repeated here for convenience as (35) and (37), while (30) is replaced by (36) as an example of the nominative-cause construction:

(35) **him ofhreow** þæs mannes (Ælfric, *Hom.* Th. 1, 192, 16)

(36) þa ofhreow **þam munece** þæs hreoflian mægenleast

then caused-pity(3-SG) the monk(DAT-SG) the leper's feebleness(NOM-SG)

'then the monk pitied the leper's feebleness' (Ælfric, *Hom.* Th. 1, 336, 10)

(37) þa **se mæssepreost** þæs mannes ofhreow (Ælfric, *Hom.* 2, 142, 262)

As previously discussed, the examples illustrate three different patterns with regard to case marking. The experiencer is marked by dative case and the cause argument by genitive case in (35), which is replaced by nominative case marking of the cause argument in (36) and by nominative case marking of the experiencer argument in (37). The observable distribution of case marking is crucial to Fischer/van der Leek's analysis, which, unlike Elmer's account, takes instances with two nominal complements as its starting point, while clausal complements are only integrated later as an optional realisation of the cause argument, and expletive constituents are largely disregarded, since they do not constitute arguments of the verb. Their proposed lexical entry for the subcategorisation of impersonal verbs thus contains a nominal experiencer marked by dative case, while the role of cause is assigned either to a nominal complement marked by genitive case or to a clausal complement. In this form, their lexical entry only accounts for the "true" impersonal type *i* represented by (35) above, however. The apparent variation of this morphological pattern with patterns like those in (36) and (37) seems difficult to account for under the assumption of syntax as a self-contained system without recourse to semantic or pragmatic factors (on a similar note cf. Denison 1990: 130, 1993: 96). An elegant solution to the problem is, in fact, proposed by Fischer/van der Leek (1983: 357f.), but before their proposal is discussed, some additional issues of variation need to be addressed.

2.1.4.1.2. Additional Surface Variation

2.1.4.1.2.1. Accusative Experiencers

The fact that the experiencer argument of impersonal verbs does not necessarily take the dative but can also be marked by accusative case is already apparent from the initial example given in

(1) above. Regarding the distribution of the two cases, some lexical preferences are discernible. Accusative occurs regularly or, at least, predominantly with *þyrstan* ‘to thirst’, *langian* ‘to cause longing’, *lystan* ‘to cause pleasure’ and *sceamian* ‘to feel ashamed’, while *hreowan* ‘to rue’, *mætan* ‘to dream’ and also *hyngrian* ‘to hunger’ are constructed with either dative or accusative experiencers (cf. the overview in Mitchell 1985: 428f.). Examples of such verbs with either regular or optional accusative experiencers are given in (38) and (39):

(38) oft **þone** **geþyldegestan** scamaþ þæs siges

often the most-patient(ACC-SG) causes-shame the victory(GEN-SG)

‘the most patient one is often ashamed of his victory’ (Alfred, *C. P.* 227, 19)

(39) hreaw **hine** swiþe þæt he [...]

rued(3-SG) him(ACC-SG) very that he

‘it grieved him very much that he’ (*Genesis* 1276)

A complicating factor lies the fact that the grammatical distinction between accusative and dative is not always morphologically marked in Old English, which means that many instances of experiencer arguments are, in fact, ambiguous in this regard (cf. Visser 1963: 20). For nouns in the singular, the distinction hinges upon the preservation of the dative ending *-e* in the most common declension of masculine and neuter nouns (cf. e.g. acc. sg. *god* vs. dat. sg. *god-e*), while feminine nouns generally require the presence of a demonstrative pronoun or an attributive adjective for disambiguation (cf. e.g. the demonstrative acc. sg. *þa* vs. dat.-gen. sg. *þære*). In the case of pronouns, the third-person pronouns exhibit distinct forms in the accusative and dative (cf. acc. sg. *hine* vs. dat. sg. *him*, nom.-acc. pl. *hī* vs. dat. pl. *him*), while first- and second-person pronouns exhibit distinct accusative forms only in early or poetic texts (cf. e.g. 1. sg. *mec* and pl. *usic* vs. later acc.-dat. *me* and acc.-dat. *us*, while the dual only has indistinctive acc.-dat. *unc*).

It is interesting to note that cognate verbs in related languages frequently exhibit the same kind of case marking, cf. e.g. ON *langa* and OHG *langēn*, which govern the accusative like OE *langian*, and ON *þykja*, which governs the dative like OE *þyncan*, although OHG *dunchen* occurs with either dative or accusative case. An historical explanation for the difference in case government between different verbs, but also for the occurrence of different cases with a single verb would thus be desirable. Allen (1995: 25) assumes that a semantic distinction was originally present, but that the different cases had already been lexicalised by

the OE period. The traces of a semantic distinction are typically characterised as a presumably more direct affectedness of experiencers marked by accusative case, which would have to include *þyrstan* as well as *langian* and verbs of possession such as ON *vanta* ‘to lack’ (cf. ME *wanten* ‘id.’). Möhlig-Falke (2012: 120) notes a predominance of the accusative with physiological sensations and emotions that are optionally accompanied by a change of state, cf. e.g. the secondary sense of *sceamian* ‘to feel shame, blush’, while verbs of cognition and verbs of happenstance appear to occur more frequently with dative experiencers. The distinction is not clear-cut, however, since *hyngrian*, for example, would be expected to behave similarly to *þyrstan*, and there are also observable differences between semantic equivalents such as OE *mætan* ‘to dream’ (cf. ME *dremen*) and ON *dreyma* ‘id.’ on the one hand, which normally take the accusative, and OHG *troumen* ‘id.’ on the other, which, in its impersonal use, takes the dative case. As it stands, the suggested semantic distinction between accusative and dative experiencers in Old English appears to require at least some lexical specifications. Although dative case is clearly the more frequent realisation overall, the designation of the relevant arguments as oblique experiencers appears to be more appropriate, since it includes the alternative realisation by accusative case according to the definition applied here. The main point, however, is that this type of variation between different impersonal verbs and within a given verb should be accounted for in a subcategorisation frame that is intended to provide a universal analysis of impersonal verbs.

2.1.4.1.2.2. Non-Obligatory Arguments

One reason for the relative frequency of dative case marking with impersonal verbs is the theoretical inclusion of lexical items whose relevant argument is typically labelled as experiencer, although its semantic role is, perhaps, better described as that of a beneficiary or by an alternative term as dative of interest (cf. the Latin *dativus commodi / incommodi*). The role of beneficiary is construed as metaphorical possession by Traugott (1992: 206), and it is thus similar to that of recipient. Constituents bearing the semantic role of beneficiary are consistently marked by dative case, but they do not qualify as verbal arguments by the same criteria as experiencer arguments, since the specification of an affected participant is usually not obligatory in these cases. As a result, verbs of happenstance such as *gelimpan* can occur without a relevant dative argument, a fact that was already illustrated in (8) above, repeated below as (40), and similar facts are true for *behove*-verbs and *seem*-verbs exemplified in (41) and (42) respectively:

- (40) þa gelamp hit þæt æt þam gyftum [...] (Ælfric, *Hom.* Th. 1, 58, 12)
- (41) now it behoueth to make mention of another order (Shute, *Archit.* 12v)
- (42) his grisliche teþ senden of swart irn
 ‘his horrible teeth seemed [to be] of black iron’ (*Marh.* 20, 24)

The potential argument status of the missing constituent can be tested with the help of two criteria suggested by Comrie (1993: 907), viz. the analogy with obligatory constituents in parallel structures, and variation of the constituent’s meaning as a function of the predicate meaning. With regard to the former, the analogy of impersonal verbs expressing physiological states and emotions seems to support the analysis of the missing dative constituent as a verbal argument, although a more specific definition of “parallel structures” would be desirable. With regard to the latter criterion, it seems obvious that the capacity for omission and a relatively independent constituent meaning appear to be related phenomena. While the examples in (41) and (42) do not imply the actual lack of a relevant participant, its omission shows that the event designated by the predicate meaning has a non-specific influence on the experiencer. According to Elmer (1981: 133), from whom the example in (42) is taken, a “generic” experiencer can be inferred in this instance, and a general kind of obligation is also expressed in (41), although it primarily pertains to the author in the given context. The clearest example of optionality, however, is that of *gelimpan* in (40). The full passage describes the Wedding at Cana, where Jesus, according to Christian tradition, turned water into wine in order to replenish the exhausted supplies. A dative constituent expressing the affected participants could easily have been supplied (e.g. *þa gelamp gestum* ‘then it happened to the guests’), but the event takes place regardless of whether or not the affected participants are specified. A physiological or emotional sensation such as *hyngrian* or *hreowan*, on the other hand, logically presupposes an experiencer, whose specification is therefore inherently more likely. In terms of obligatoriness, there is thus a notable difference between experiencer verbs in the more literal sense and verbs of happenstance or, to a lesser degree, verbs of obligation and constructions with *seem* in a wider sense. Incidentally, this also demonstrates that the variation between dative and accusative case marking is much more pronounced among actual experiencer verbs, which seems to accord with the fact that the kind of direct affectedness assumed for the accusative is more readily conceivable with an obligatory experiencer argument.

Conversely, the second argument, which is regularly absent in the case of verbs denoting physiological sensations, can be optionally absent from constructions of verbs of emotion, a

phenomenon which is sometimes called an “absolute construction” (cf. e.g. Elmer 1981: 73, the term is also used by Mitchell (1985: 428)). Anderson (1988: 13) simply takes this to be an extension of the basic pattern, and one might, indeed, argue that a “generic” or non-specific cause is to be understood in such instances. It is nevertheless important to note that the argument structure of impersonal verbs is not uniformly realised across different semantic classes. Within the more narrowly defined group of two-argument verbs, verbs of happenstance as well as verbs of obligation and raising constructions with *seem* illustrate optionality of the experiencer to different degrees, while verbs of emotion may leave the second argument unspecified. This distinction can be grounded theoretically in a more specific analysis of the second argument as theme in the former case and as source (or cause) in the latter (cf. Anderson 1986: 174–176 and the discussion below). Clearly, the predicate meaning of a verb that profiles a theme argument would be more autonomous than that of a verb that profiles a source argument, since the latter usually requires the expression or at least implies the presences of an experiencer.

2.1.4.1.2.3. PP Constituents

The final point of variation to be discussed in the present context is the realisation of one of the verbal arguments by a PP. This phenomenon is more frequently observed for the second argument of impersonal verbs than for the experiencer argument, but occasional instances of PP experiencers are found as well. An illustration of the former is presented in (43) below:

(43) ȝet **hym** shulde arewen of the arrerage

yet him(OBJ-SG) would(3-SG) rue(INF) of the arrears

‘yet he would regret the arrears’ (*Pol. Songs*, 138)

In this example, the argument expressing the source of regret is realised by a PP *of the arrerage* ‘of the arrears’, which is headed by the preposition *of*. Besides this, PPs headed by *on* are attested for the second argument of ME *reuen*. The general replacement of morphological case by the use prepositions is usually considered to be a ME development (cf. e.g. Fischer/van der Leek 1983: 361 and their discussion), but there are examples already in Old English of the prepositional use of OE *langian*, which occurs in constructions with *on* and *æfter*. It is also quite common to simply regard PPs as structural equivalents to case-marked constituents (cf. e.g. Allen 1995: 70), but they nevertheless constitute a kind of variation that is not immediately accounted for by a lexical entry that is specified in terms of case marking.

2.1.4.1.3. Optionality of Lexical Case

The examples in the preceding paragraphs have illustrated that the realisation of the impersonal type i, which involves a dative experiencer and a genitive cause argument in Fischer/van der Leek's lexical entry, is, in fact, more varied than such a formalisation suggests, since it includes accusative experiencers and a PP constituent as well as optional omissions of either of the two arguments. In terms of semantic roles, an additional distinction between cause arguments and theme arguments seems appropriate in order to account for the difference between experiencer verbs in a narrower sense and verbs of happenstance and obligation. Also, if the case marking of *beawas* in (18) above is analysed as accusative, this would constitute a third grammatical option for the cause argument (thus Fischer/van der Leek 1983: 348). In view of the fact that accusative case marking of cause arguments is not unambiguously attested elsewhere, this seems less preferable, however, than the assumption of sporadic lack of verb agreement (Allen 1986a: 388) or early phonological loss of final *-n* in the verbal form *gelicade* (Ogura 1990: 36). Regardless of these issues of variation, the patterns are all analysed as impersonal according to the definition as subjectless and subsumed under a single type. The parallel occurrences of type ii and type iii with a nominative cause and a nominative experiencer argument do not conform to the same definition of impersonal constructions, however, and require additional lexical entries in Fischer/van der Leek's initial proposal. Crucially, this analysis is ultimately revised as a single lexical entry with the additional stipulation that lexical case is assigned optionally to the experiencer argument and the cause argument of impersonal verbs, while lexical case is assumed to be assigned obligatorily in the case of non-impersonal verbs (cf. Fischer/van der Leek 1983: 357f.). The conditions under which lexical case assignment fails to apply are not quite clear, however. NP movement into subject position is adduced to account for instances with an initial nominative cause (Elmer's type Ia), which is said to receive structural case governed directly by tense due to its occurrence in subject position. Whether or not NP movement and the assignment of structural case are related phenomena is not elaborated upon, although this would be an interesting point, since a fronted cause argument could, conceivably, remain in the genitive case. More problematic, however, are instances with a nominative cause argument in postverbal position (Elmer's type Ib). It receives structural nominative case despite not being in subject position, a fact which Fischer/van der Leek propose to explain with the concept of chain-government developed by van Besten (1981, apud Fischer/van der Leek 1983: 358–360). This becomes operative when, “for whatever reason”, an argument does not receive case from its governing verb, in which case structural nominative is chain-governed by tense. Again, no particular reason for the failure of case government by the verb is specified, and the

conspicuous restriction of the proposed optionality of lexical case government to experiencer verbs is noted (Fischer/van der Leek 1983: 357¹¹) but not discussed any further. Most importantly, though, from the point of view taken in the present investigation, their formal account does not integrate semantic considerations regarding the preference of a particular construction type over another. It thus constitutes an expedient, albeit somewhat simplified, way of modelling the constructional patterns observed for experiencer verbs, but these are stated in purely formal terms (for Fischer/van der Leek's functional account, cf. the discussion in section 2.1.4.4 below).

2.1.4.2. Case Grammar

Similar problems are present in Anderson's analysis (1984, 1986, 1988), which is based in the general framework of Case Grammar. This theory operates with so-called "case relations", the theoretical equivalent of "semantic roles" or "θ-roles", which are regarded as the only necessary specifications in lexical subcategorisation frames, while the assignment of grammatical relations like subject and object is considered to be derivative (cf. Anderson 1986: 176, 1988: 1–4). Specifically, Anderson assumes case relations to be of a compound nature (cf. Anderson 1984: 255–257). Experiencer arguments, for example, are interpreted as compounds of the basic relations ergative and locative, since they arguably share certain properties with both simple ergatives and simple locatives. This observation seems to be related to the uneven distribution of behavioural subject properties that was discussed earlier, which might lend support to the notion of gradient subjecthood. According to Anderson, however, subjecthood is a discrete property that is predictably assigned to the argument that bears the highest-ranking case relation according to the hierarchy in (44), which is taken from Anderson (1988: 2):

(44) $\text{erg} > \text{erg} + \text{case} > \text{abs} + \text{case} > \text{abs}$

This hierarchy includes only the most basic case relations. The ergative (*erg*) refers to agentive participants, while the absolutive (*abs*) refers to (neutral) theme arguments and affected entities. Other case relations (*case*) are restricted to locative (*loc*) and ablative (*abl*) by Anderson (1988: 3), the latter of which can be considered the equivalent to a source argument or cause. The stipulation that subjecthood is assigned to the leftmost argument in the above hierarchy predicts the nominative-experiencer construction of type iii (Anderson's example (51b), cf. 1988: 15, also 1986: 171f.), in which the experiencer argument, bearing the compound case relation ergative + locative, is selected as subject in the absence of a simple ergative argument. Type ii

with a nominative cause should not occur, however, since the experiencer outranks the cause argument, which bears the lowest case relation ablative. In order to address this problem, Anderson (1988: 15) proposes an analysis of the nominative cause argument in type ii as ergative, while it would remain ablative in types i and iii. Although this analysis appears to be circular in that the assignment of nominative case to an argument thus marked as subject is taken as an indication of its case relation, while case relations are otherwise regarded as the basis of a derivative process of the assignment of subjecthood, it does convey the impression discussed earlier that cause arguments share certain properties with agents. Under this view, it would be understandable that both the experiencer and the cause argument can occur as subjects in the absence of a typical agent, which would be more prototypically associated with the subject relation. Within the formal account, however, neither the ambivalent realisation of the cause argument as ablative or ergative, nor the optional assignment of subjecthood, which constitutes an additional requirement due to the absence of nominative constituents in the impersonal type i, are meaningfully explained. The seemingly parallel occurrence of types i–iii thus proves difficult to account for by a single synchronic analysis also in terms of Case Grammar.

2.1.4.3. Problems of a Universal Account

2.1.4.3.1. Restricted Occurrence of *Lician*

Apart from its failure to meaningfully account for the optionality of lexical case marking (or “subject formation”, in Anderson’s terms), the validity of Fischer/van der Leek’s universal treatment of impersonal verbs has also been called into question on empirical grounds, cf. especially Denison 1990: 113–118. The verb *lician*, for example, is not recorded by Elmer in the impersonal type i, nor in type iii with a nominative experiencer (except in so-called absolute constructions without a second argument), but only in type ii with a nominative cause (Elmer’s type I), or in constructions with a clausal complement (Elmer’s type S), in which latter case the classification as type i or type ii is ambiguous. Regarding the potential evidence of the other types, the following observations can be made. The existence of type i with *lician* mainly depends on the interpretation of examples like (18) above, in which *peawas* is ambiguously marked for nominative or accusative case. Denison (1990: 114) reports four additional examples of potential cause arguments of *lician* in the accusative, but he does not offer a full discussion. If the examples are valid, type-i constructions of *lician* can be acknowledged to occur in a special variety with accusative case, which does not seem to be attested elsewhere,

however, and which would be rare overall. As to type iii, its occurrence with *lician* in OE is generally not disputed, despite the apparent gap in Elmer's material. The question is rather whether the relevant examples are to be considered manifestations of indigenous OE syntax, or whether they are the result of close translations from Latin, which would render their significance for the analysis of Old English suspect.

The former view is held by Waltz (2000: 284f.), who considers the variation between different constructional types a function of information structure. Nominative experiencers are assumed to express unimportant background information, while nominative cause arguments are assumed to place special emphasis on important events. Waltz also points out that a personal verb in a Latin exemplar may be rendered by an impersonal verb in Old English (cf. e.g. the translation of Lat. *rubesco* as *me scamap*, cited by Delbrück 1900: 33f. apud Waltz 2000: 286), which seems to point to the relative independence of the OE translation. Regardless of whether or not her interpretation of the functional difference in terms of information structure is correct, however, its establishment on the basis of individual interpretations of relatively sparse data cannot fully dispel concerns about a circular argument. Different renditions of *Mt. 15, 32 misereor* 'I commiserate', for example, which involve either a nominative experiencer *ic gemiltsige* or an oblique experiencer *mec hreoweþ*, are interpreted with reference to the scribe's perceived attempt at glorifying and emphasising the feeling of compassion in the latter case (cf. Waltz 2000: 298), but the only indication of this appears to be the use of an oblique experiencer itself. In the case of *lician*, the gospel according to Mark and the gospel according to Luke present two different versions of the Latin original, the former of which is quoted below as (45) alongside its OE translation, both taken from Waltz (2000: 284):

(45) tu es filius meus dilectus in te complacui

þu arþ sunu min leaf on þec ic wel licade

you are son my dear on you I well liked

'you are my dear son, in you I took great pleasure' (*OE Gosp. Mk. (Li) 1, 11*)

Both the Latin and the OE version are constructed personally with a verb in the first person singular, in addition to which the OE translation contains the nominative pronoun *ic*. The variant reading *in te complacuit mihi* in *Luk. 3, 22*, on the other hand, contains a dative experiencer, which is translated into Old English as *on þec licap me* (Li, Ru) and *on þe me gelicode* (WSCp) with an oblique experiencer *me*. To account for the difference between the two gospels, Waltz (2000: 301) invokes a special emphasis on the event of pride expressed at

the baptism of Jesus in Luke, while, in Mark, the event is said to be presented in terms of a general truth. The concurrence of the OE and the Latin version in both cases is supposed to reflect the fact that both authors happened to agree in their view of the situation. This reasoning is hardly convincing, however, and the passage rather provides an example of the potential dependence of a translation upon its exemplar.

Incidentally, the translations cited above also differ from the supposed basic pattern of impersonal verbs by marking the cause argument as a PP rather than by genitive case, which, again, accords with the Latin example. Denison (1990: 115) proposes to view such PPs as peripheral to the predicate rather than as verbal arguments, similar to PDE constructions such as *I did not like this in him*. In absence of a proper cause argument, this assumption does not seem to apply to *lician* in (45), however. While PP complements do occur more widely with certain verbs in Middle English, their use in the given examples appears to be modelled after the Latin original. Their classification as instances of Latinate syntax is extended to the personal construction of *lician* with a nominative experiencer by Allen (1986a: 387), who argues for the general exclusion of interlinear translations and word-for-word glosses from investigations of Old English in order to exclude as much interference of Latin as possible (cf. Allen 1997: 18⁴). A similar point is made by Denison (1990: 114–116), although he is more reluctant to dismiss the relevant examples as evidence, since they arguably require some degree of closeness to the indigenous syntax in order to be acceptable for the translation. In any case, it seems that the constructional types i–iii were not simply in free variation in Old English with the verb *lician*, for which type ii with an additional nominative constituent or the ambiguous type i/ii with a clausal complement (on this designation cf. Denison 1990: 14) were clearly dominant.

2.1.4.3.2. Other Verbs

Apart from the potential variation with *lician*, evidence of the occurrence of nominative experiencers besides oblique experiencers during the OE period has been adduced with reference to a range of other verbs. One piece of evidence comes from the verbs *hyngrian* and *byrstan*, which are typical one-argument verbs (cf. (46)), but which can also be constructed with an additional cause argument in their figurative sense (cf. (47)):

(46) eadige synd ge þe hingriap nu

blessed are you(NOM-PL) who(PTCL) hunger(3-PL) now

‘blessed are they who are hungry now’ (*OE Gosp. Lk. 6, 21*)

(47) **min sawl** on þe swyþe þyrstep

my soul(NOM-SG) after you very thirsts(3-SG)

‘my soul longs very much for you’ (*Psalm* Th. 62, 1)

Both instances can be analysed as personal constructions, but in (46), this interpretation relies on the plural ending *-iap*, which contrasts with the third-person singular ending *-ap*, while the realisation of the experiencer argument by the relative particle *þe* itself is morphologically ambiguous. In (47), nominative case marking can be directly inferred from the possessive determiner *min* ‘my’. Another verb that exhibits variable case marking of the experiencer is OE *behofian*, which starts out as a typical type-iii verb with a nominative experiencer (cf. Allen 1997: 4, also Loureiro-Porto 2009: 109–140) and eventually acquires the construction with an oblique experiencer in Early Middle English. In addition to these verbs, Fischer/van der Leek (1983: 352–354) note the optional occurrence of nominative experiencers with OE *eglian* ‘to ail’ and OE *lystan* ‘to desire’, and a preference of nominative experiencers with OE *reccan* ‘to care’, *sceamian* ‘to feel ashamed’, OE *tweogan* ‘to doubt’ and OE *swefnian* ‘to dream’. A special case, according to the authors, is OE *þyncan* ‘to seem’, whose only potential instances with an (implicit) experiencer subject concern constructions like the one in (48), their (17):

(48) *se leoma* [...] was swiþe lang gepuht

the light(NOM-SG) ... was / has very long thought /seemed

‘the light was considered / has seemed very long’ (*Chron.* Pl. 233)

Fischer/van der Leek (1983: 354) analyse this instance as a passive construction, in which *þyncan* exhibits the “receptive” meaning associated with a nominative experiencer. An analogous distinction is made between “receptive” *lician*, which is glossed as ‘like’, and “causative” *lician*, which is glossed as ‘please’. The verb *þyncan* is accordingly translated as ‘consider’ in the above example. Anderson (1988: 22) points out, however, that examples like (48) can also be regarded as instances of a tense-aspect construction with the auxiliary *wesan*, whose use instead of the alternative *habban* ‘have’ would be regular for intransitive verbs (cf. also Denison 1990: 116f.). Under this analysis, the example simply constitutes an instance of a type-ii construction with a nominative cause and no experiencer, similar to (42) above.

The synopsis of constructional patterns supplied by Elmer (1981: 63) indicates that not only the attestation of type iii is uncertain for *þyncan* at best, but also that diagnostic instances of type i with a genitive complement appear to be absent for this verb. Similar observations

regarding the absence of type i outside of ambiguous instances with a finite or non-finite complement clause apply to *happen*-verbs and *behove*-verbs, which are thus set apart from experiencer verbs proper even further in that not only their expression of a specific experiencer argument is non-obligatory and restricted to constituents marked by dative (or nominative) case, but also their second argument is generally not realised by a genitive complement (or PP), even though this is what the proposal of a lexical entry in which the realisation of the second argument by a complement clause is merely specified as an optional alternative suggests. This disparity between the complementation patterns observed for different lexical classes demonstrates, once more, that the common theoretical conflation of clausal and nominal complements expediently conceals the limited applicability of a universal account of impersonal verbs that regards type N (or type i) as the prototypical form. Fischer/van der Leek's proposal of a single lexical entry thus requires some qualification. For *lician*, early evidence of the constructional types i and iii is very limited and largely restricted to isolated instances of potential accusative complements for type i and instances of Latinate syntax for type iii, while the constructional type ii is clearly dominant in Old English in all cases that do not involve morphologically ambiguously complement clauses. For *byncan*, the only evidence of type iii comes from potential passive constructions, while active clauses equally involve clausal complements that cannot actually vary between type i and type ii. The same is generally true for *happen*-verbs and *behove*-verbs, which do not exhibit such variation regarding the second complement, and which show a typical preference for either dative and nominative experiencers in Old English. Also the other verbs that were mentioned above but not discussed in detail can be regarded as exhibiting a preference for either oblique or nominative experiencers rather than free variation (for a discussion of the evidence cf. Allen 1995: 80–82). A uniform postulation of optional lexical case marking will thus always eclipse some form of lexical specification. The attestation of different constructional patterns with individual lexemes is an important observation that is also at the core of Möhlig-Falke's investigation of the interaction of lexical and constructional meaning. It would be misleading, however, to assume that impersonal and personal constructions varied freely and to the same extent with all of the verbs that are usually described as impersonal. The assumption of the present account will thus be that dominant constructional patterns can be established for each of the verbs under discussion during a given reference period. The dominance of a particular pattern may be purely lexicalised, but it appears to correspond at least partly to verbal semantics, since verbs of happenstance and verbs of obligation are generally excluded from type-i constructions with two nominal complements, and verbs of obligation generally appear to acquire the impersonal type iii at a relatively late

stage. At the same time, the existence of some degree of synchronic variation needs to be acknowledged for the OE period, and the question arises whether this variation can be explained in terms of different semantics or communicative functions.

2.1.4.4. Functional Aspects

As a final point in the discussion of synchronic accounts of impersonal verbs, the functional aspect of a synchronic variation between three constructional types will be considered once more. Since parallel instances of the constructional patterns appear to occur with certain verbs already in Old English, the question arises whether these exhibit demonstrably different functions. An attempt at relating the constructional difference between oblique experiencer arguments in type i and type ii and nominative experiencer arguments in type iii to a difference between foreground and background information was already encountered in the discussion of Waltz's proposal above. Reliable conclusions about information structure would probably require a much broader database than the individual examples discussed by her. Also, their possible dependence on Latin exemplars would have to be controlled for. The principal notion that a nominative cause argument may serve to place special emphasis on the causative effect compared to the use of nominative experiencer argument appears to be a plausible assumption, however. Since the morphological properties of subjects are prototypically assigned to agentive participants whenever these are present, a causative interpretation of a given event would arguably be more adequately expressed by the assignment of nominative case to the cause argument. In a similar vein, Fischer/van der Leek propose to explain the difference between type-i and type-ii constructions in terms of different degrees of transitivity (Fischer/van der Leek 1983: 350f.) or in terms of the absence or presence of participants with argument status (Fischer/van der Leek 1987: 111). They share the assumption that the constructional type ii with a nominative cause expresses a higher degree of affectedness on the part of the experiencer, but this is contrasted with type-i constructions, in which no nominative constituent is present, rather than with type-iii constructions with a nominative experiencer. In synthesis of such earlier accounts, Möhlig-Falke (2012: 151) describes the function of type-i constructions of experiencer verbs as "foregrounding the human endpoint of the transitive event" while "suppression, or backgrounding, of the initiator" is said to lead to "defocusing of the inherently dynamic nature of the involved mental process". While this characterisation obviously only applies to impersonal constructions defined by the lack of a nominative constituent, the implicit

assumption of an association of nominative case with a more active or direct involvement of the cause argument is common to all of the accounts discussed so far.

There are two points to address in relation to the above assumption. First, the variation of individual verbs between the constructional types i and ii appears to have been rather limited if the evidence of morphologically ambiguous clausal complements is disregarded. The question thus arises whether the functional analysis of impersonal constructions without a nominative constituent can be extended to constructions with clausal complements, even though the preverbal topic position of the experiencer argument, which is taken as an indication of foregrounding by Möhlig-Falke (2012: 151), is not actually distinctive in these cases. Regarding constructions with two nominal complements, Elmer (1981: 70f.) considers instances of type i (his type N) and type ii (his type I) to be lexically distributed, since the two types rarely contrast with a single verb (only *hreowan* and *offbyncan* in his collection of examples), and the examples of variation that do exist do not seem to show a clear semantic contrast. Second, Elmer (1981: 60, 69f.) actually considers type i rather than type ii the typical representative of what he calls the “causative relation”, a term adopted from Visser (1963: 355), while he considers the relation between the verb and the second argument in type ii to be of a more neutral kind. The reason for this seems to be that the second argument of verbs that occur in type-i constructions can generally be analysed as the source or cause of an experience, while the second argument of verbs in type-ii constructions may also represent a theme. The latter analysis arguably applies to the second argument of verbs of happenstance (“existential experiences”) and verbs of obligation (“possession”). When realised by nominal constituents, theme arguments are generally marked by nominative case, however, even if this does not coincide with an agentive interpretation. Theme arguments, in fact, appear to have a very limited potential for variable interpretations in terms of agency. This would agree with the apparent absence of variation between type-i and type-ii constructions when theme arguments are involved, but the fact that nominative case is assigned to them instead does not appear to be an indication of agentivity but rather connected to the non-obligatoriness of the experiencer argument, which makes the theme the primary participant.

A semantic distinction is equally conceivable for different forms of the experiencer argument. While van der Gaaf (1904: 90f.) remarks in the context of his investigation of textual variants involving personal *ich þenche* ‘I think’ (cf. German *denken* ‘id.’) and impersonal *me þunchep* ‘me thinks’ (cf. German *dünchen* ‘id.’) that these were apparently felt to be, to some extent, synonymous, the nominative experiencer argument of *þencan* arguably involves a more active cogniser rather than a passively affected experiencer. The variation between oblique and

nominative experiencer arguments may thus be related to perceived degrees of agentivity even in those cases in which the second argument is not marked by morphological case or not subject to the same kind of variable semantic interpretations assumed for cause arguments. Based on similar theoretical assumptions, Fischer/van der Leek (1983: 351) make a distinction between type-i and type-iii constructions in terms of volition, stating that, in type iii, “the animate experiencer is nominative subject and therefore the initiator of the ‘action’, fully involved in what the verb expresses”, whereas in type i, “the experiencer [...] is only passively related to what is expressed in the verb”. The notion of unvolitional involvement of the experiencer is expounded in more detail by McCawley (1976: 194–197), who draws on the comparison of cross-linguistic parallels with impersonal constructions like the “fake reflexive” in Russian (e.g. *mne vspomnilas’ eta noc’*, lit. ‘the night remembered itself to me’) and the “fake passive” in Japanese (e.g. *watashi ni wa soo omowareru* ‘it seems so to me’), which seem to indicate a similar kind of passive involvement of the experiencer.

The distinction between more or less agentive experiencers or a more or less volitional involvement of the experiencer obviously constitutes a matter of degree, since the verbs in question do not profile prototypical agents. The question is thus how different interpretations of the experiencer can be manifested in the data beyond theoretical preconceptions about the supposed implications of different morphological cases. Elmer (1981: 74), for example, is reluctant to distinguish between types i and iii in terms of volition, since he considers it difficult to judge whether such interpretations actually applied in the OE examples. While it can be difficult to establish such semantic nuances in historical records based purely on an interpretation of the context of individual occurrences, and while the interchangeability of cases in different textual variants appears to point to some degree of semantic equivalence, there are relatively objective semantic properties like animacy, which arguably have a bearing on the potential of a given participant to be interpreted as agentive. Since the capacity of undergoing experiences is logically restricted to animate participants, this potential is generally present with experiencer arguments of impersonal verbs, but it is a variable property of the cause argument of verbs of emotion, which may be either animate or inanimate. Theme arguments, on the other hand, are generally marked by nominative case, regardless of their potential for an agentive interpretation. This seems to be related to the fact that the experiencer is not presupposed but largely optional with the relevant verbs, which makes the theme the primary participant in conceptual perspective. A more detailed discussion of the semantic roles associated with impersonal verbs thus seems to be required and will be provided in section 2.2.2.1.3 of the theoretical background. At this point, it should suffice to note that both the experiencer

argument of verbs of emotion and its corresponding cause argument, for which the more neutral term “stimulus argument” will be adopted in the remainder of the text, have the potential for a more or less agentive interpretation, which may be linked to their variable case marking by either oblique or nominative case. The integration of this semantic distinction into a diachronic account of the development of impersonal verbs will be the main focus of the subsequent sections.

2.2. Diachronic Change

There are numerous theoretical approaches to the general phenomenon of language change, which take a range of different perspectives and are aimed at various degrees of comprehensiveness (for some of the more relevant frameworks in the present context cf. Harris/Campbell 1995, Croft 2000, Hopper/Traugott 2003, Roberts 2007). An exhaustive survey of the different approaches would be beyond the scope of the present discussion, but there are a number of points that are of particular relevance to the development of impersonal verbs in English. A fundamental issue that has been raised is the question of whether a dedicated framework of diachronic change constitutes a theoretical requirement, or whether a restrictive account of synchronic grammar is sufficient to account for the observed changes in language. The latter view is repeatedly expressed by Lightfoot (1979: viii, 35, 42, 343 et passim), who essentially regards historically attested changes as a testing ground for synchronic theories of language. The requirement of a synchronic framework for the investigation of diachronic syntax in particular is also stressed by Fischer (2007: 14–17), which is certainly a valid point, since the basic units that are subject to diachronic change in syntax are not as easily defined and certainly exhibit a different level of complexity than the equivalent units in phonology and morphology. A major problem is that the configurational rules underlying the observable attestations of linguistic competence can be formalised in a variety of different ways. These formalisations can, of course, be compared to each other in terms of their economy and adequacy of description, but whether or not they constitute a mental reality or merely convenient descriptive tools is a different question. The very object of syntactic change thus constitutes a matter of theoretical debate, and synchronic and diachronic frameworks will always be equally bound to a similar set of theoretical preconceptions.

Whether the intrinsic properties of a specific theoretical analysis can ever be formulated in a way that is restrictive enough to actually account for observable changes remains highly doubtful, however. The basic question in this context is whether language-internal factors can

be established as the primary cause of a given diachronic change or whether external factors such as social identification and language contact generally take precedence over structural considerations. The former view immediately leads to the apparent paradox of how linguistic structures that entail the apparent need of a subsequent therapeutic change could even arise, but it is at the heart of one of the most frequently observed proposals regarding the development of impersonal verbs. According to this proposal, the changes in one linguistic domain lead to complications in another domain, which eventually requires a large-scale revision of the linguistic system that takes place in the form of an observable cataclysmic change. This is, of course, Lightfoot's original proposal of change by reanalysis (1979 and later revisions), which will be considered in more detail below. A second important notion underlying this theory is the idea that language change crucially depends on the scenario of first-language acquisition. Within the framework of transformational grammar, it is quite reasonable to assume that the surface structures that constitute the priming experience of a language learner can lead to the abstraction of a mental grammar that differs from the one (or the ones) that produced the input, which leads to linguistic change by a process called abduction (cf. Andersen 1973: 767). The conceptualisation of a speech community's shared linguistic competence as a set of distinct mental realities obviously leads to other questions like how the converging tendencies in long-term diachronic changes are to be explained.

A universal tendency towards the establishment of syntactic functions is assumed for the development of impersonal verbs by von Steffens-Montag (1983), although this effect is taken to be fully realised only in English, while the other languages in her study present stages of this development that differ in several ways. A related question in this context is the extent to which linguistic changes may depend on or be triggered by structural ambiguity resulting from a lack of morphological distinctions. This notion constitutes a basic assumption in the version of reanalysis theory advanced by Jespersen in the third part of his *Modern English Grammar* (1927), although it is not crucial to Lightfoot's later reformulation of this theory in transformational terms, in which he assumes the applicability of syntactic reanalysis to constituents marked by oblique case. An obvious problem for any account involving morphological ambiguity is, of course, the fact that morphological case marking may be distributed unevenly across the inflectional classes of a language, and the question of how much syncretism constitutes a degree of surface ambiguity that is sufficient to trigger syntactic changes is one that has no definite answer (cf. also Anderson 2001b: 234). A criticism that has been levelled against the notion that morphological ambiguity provided the basis for a reanalysis of oblique experiencer arguments of impersonal verbs as subjects is the fact that at

least some morphological distinctions are retained in the pronominal system of English, which are arguably sufficient to disambiguate any occurrences of morphologically ambiguous constituents, especially since unambiguous pronouns are considered to be the more typical realisation of experiencer arguments. It is, of course, conceivable that also a limited number of morphologically ambiguous cases can be sufficient in triggering a reanalysis that is later extended to further contexts. This assumption is held by Harris/Campbell (1995: 397, cf. also 70–72), whose diachronic framework involves reanalysis and extension as two major factors in diachronic syntactic change. The more widely held assumption, however, appears to be that syntactic changes precede any simplifications in related areas of the morphological system that become redundant (cf. e.g. Fischer 2013: 137). In accordance with this view, von See Franz-Montag argues that the development of impersonal verbs needs to be regarded as generally independent of the distinctiveness of morphological case marking, pointing to comparative evidence from a range of other languages. The replacement of impersonal constructions is found to occur much later in French, for example, despite a similarly early loss of morphological distinctions in the nominal system (cf. von See Franz-Montag 1983: 239f.), while the inventory of impersonal verbs and constructions in German is said to be reduced in comparison to earlier stages of the language, even though the general case distinctions are largely maintained with the help of determiners (cf. von See Franz-Montag 1983: 172f.). Whether her assumption of a universal tendency towards syntactic relations constitutes an explanation or a description of the observed changes is, of course, another question, and while potential factors in the apparent regularities of diachronic change would require a much broader discussion, the associated observation that genetically related languages can represent different stages of the envisaged development necessarily involves the acknowledgement of a diversification that seems difficult to account for if not by reference to language-external factors.

Social and pragmatic functions as well as language contact have become more widely recognised as influential factors in language change compared to earlier accounts that rely more heavily on structural language properties. Contact-induced changes constitute the third primary mechanism recognised by Harris/Campbell (1995) in addition to reanalysis and extension, for example, and they are also one of the principal types of change distinguished by Anderson (2001a: 33f.), who contrasts evolutive changes, which are taken to spread from unmarked contexts into marked contexts, and contact-induced changes, which are taken to spread from marked contexts into unmarked contexts. The directionality of change and the definition of markedness in this context will play a role in the discussion in section 2.2.2.1.2 below, but, for now, it is sufficient to note that both of the aforementioned frameworks incorporate language-

external factors besides structural motivations into their account of diachronic change. The notion of frequency as an important factor in the process of change has been systematically discussed by Bybee (2007, cf. also 2006) and presents a more radical departure from earlier accounts in that it presupposes an understanding of language that involves a set of diachronically variable and synchronically non-discrete mental representations, which are subject not only to long-term diachronic change across generations of speakers, but also to continual speaker-internal negotiations that result from the changeable properties of language usage. The assumption of structural causes of change that are founded in a linguistic system that is acquired in a more or less rigid form during the priming experience thus takes second place to the semantic and communicative functions that linguistic structures perform.

An integration of the functional dimension of linguistic structures is equally central to diachronic construction grammar, which is based on a representational synthesis of the formal and functional dimensions of language in the form of constructions, which exhibit different levels of schematicity (cf. Barðdal/Gildea 2015 and Traugott 2015 in the same volume, on the theoretical basis cf. Goldberg 1995). The assumption that the form-meaning pairings commonly assumed for lexemes extend to the level of constructions leads to a reformulation of diachronic syntax in terms of constructional changes, which may affect the meaning aspect, the formal aspect, or the mapping of these two aspects, and which can also lead to the development of innovative constructions (cf. Barðdal/Gildea 2015: 14–16). It is also regarded as a solution to the perceived problem of comparative diachronic syntax of establishing the regular correspondences on which the reconstruction of a prehistoric stage of a language could be based. A version of the framework is applied to Indo-European (IE) constructions with an oblique subject by Barðdal/Smitherman (2013), who reconstruct such constructions for Proto-Indo-European (PIE) based on their investigation of a set of five cognate verbs. The distribution of these cognate constructions is often limited to a single branch, however, and the high degree of schematicity of the relevant constructions makes their comparative evidence less compelling than the authors appear to assume. Another crucial aspect in the reconstruction of prehistoric stages as well as the systematisation of historically attested changes is the directionality of change (cf. Harris/Campbell 1995: 344–353, 361–363 and Roberts 2007: 357–376 for discussions). This aspect has been a fundamental issue in research devoted to grammaticalisation (cf. Hopper/Traugott 2003), which is commonly defined as the process by which lexical units take on more grammatical functions. Since this process involves innovative interpretations of existing structures, which are taken to arise, at least in part, by the conventionalisation of conversational implicatures, the mechanism of reanalysis can be

regarded as an important factor in the framework of grammaticalisation as well, but the integration of semantic and pragmatic factors distinguishes it from the more formal accounts of transformational grammar, and its application to individual lexemes or syntagmata has a narrower scope than the constructional changes outlined by Barðdal/Gildea.

The spectrum of existing theoretical orientations that can and have been adopted in the diachronic treatment of impersonal verbs is, of course, much wider than the present discussion is able to represent, and the superficial survey of some of the more recent theoretical tendencies leaves much to be desired in terms of detail and comparative evaluation. The central question in the present context, however, is which of the aforementioned frameworks lends itself most suitably to a diachronic account of impersonal verbs. Since verbal argument structure can be conceptualised in terms of more schematic constructions, diachronic construction grammar seems to be a promising candidate, even though the reconstruction of a putative prehistoric origin of the impersonal construction cannot be profitably expanded by the present account. The inclusion of a functional aspect, which is here taken to involve both semantic and communicative functions, can be regarded as an essential improvement of earlier formal accounts, since it reflects the notion that syntactic configurations are not restricted to concerns of grammaticality according to an autonomous system of rule-based derivations or other formalised systems, but instead involved in the essential function of language as a tool of communication. Möhlig-Falke's synchronic characterisation of impersonal constructions as foregrounding the human experiencer and backgrounding the initiator takes this functional dimension into account, although it strictly speaking only applies to constructions without a nominative constituent or to constructions with a morphologically ambiguous clausal complement. A potential additional question is, then, whether the constructional meaning assumed for Old English can be demonstrated to have undergone an independent diachronic change in the history of English.

A second question relates to the formal type of constructional changes distinguished by Barðdal/Gildea (2015). In the present case, the observed changes in formal aspects of the construction can only be regarded as a single process in terms of an idealisation of several more disparate processes, while the single aspect of oblique case marking of the experiencer, which the present study takes to be the main definitional criterion of impersonal constructions, essentially appears to imply a loss of the impersonal construction and its attraction into the more productive pattern of personal constructions. A systematic account of how this formal change unfolds appears to benefit from references to syntactic units below the construction level, however, which, again, require at least some degree of formalisation. While neither the

framework of diachronic construction grammar nor any other restrictive theoretical framework will be explicitly adopted in the present account, the basic assumption that the formal units of any syntactic analysis need to be considered in relation to their semantic and communicative function will be prevalent. The fact that a theory that is based in a derivational formalisation of syntax provides the starting point of the account below is an immediate consequence of the influence this orientation had on earlier discussions of impersonal verbs. It should not be taken to imply that a representation of syntax as an autonomous system of discrete units, whose mental reality is not even assumed by some of its adherents, is given preference over a usage-based conceptualisation of language as an evolutive and domain-general tool for communication. This also seems less problematic in view of the fact that the associated mechanisms of reanalysis and actualisation persist in some form or another in more recent theoretical frameworks like diachronic construction grammar and grammaticalisation, which seems to suggest that they constitute fundamental mechanisms of language change that are recognised across theoretical boundaries. The first part of the remainder of this section will thus provide a more detailed introduction to reanalysis theory within earlier versions of transformational grammar as well as to some of its most important criticisms (section 2.2.1.1), which will be compared to Allen's diachronic account of impersonal verbs (1995), which is set in the framework of Lexical Functional Grammar (section 2.2.1.2), and to later revisions in terms Principles and Parameters (section 2.2.1.3). This comparison also affords a discussion of the relative benefits of theory-driven and empirical research, the latter of which has been greatly facilitated over the last decades by the compilation of electronic corpora. The present study aims at making a contribution to this strand of research by testing a semantic hypothesis that is derived from the application of the theoretical notion of actualisation to the development of impersonal verbs. The discussion of this notion and further basic theoretical concepts, including markedness and semantic roles, will be followed by the formulation of the hypothesis and its operationalisation for the purposes of the corpus study (sections 2.2.2.1–2.2.2.3).

2.2.1. Earlier Accounts

2.2.1.1. Reanalysis Theory

2.2.1.1.1. Outline

The idea that the loss of impersonal constructions was triggered by a mechanism of reanalysis was first expressed by Jespersen (1894: 216) and supported in a more detailed study by van der

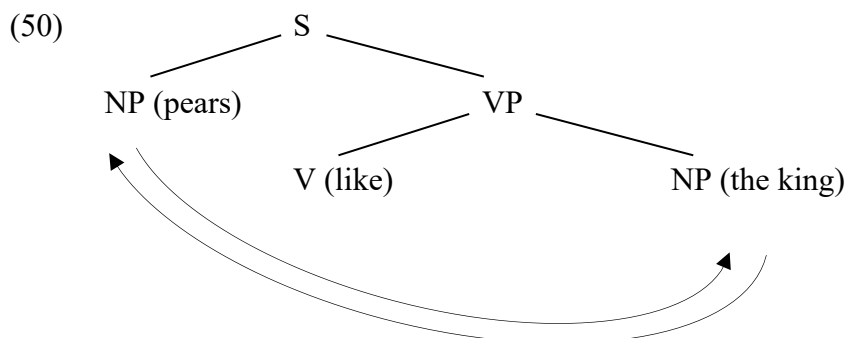
Gaaf (1904: 2). Although both scholars noted early occurrences of constructions with nominative experiencers besides constructions with oblique experiencers (cf. Jespersen 1894: 217, van der Gaaf 1904: 155–162), they essentially regarded the development of impersonal verbs as one of diachronic replacement. Focusing on two-argument verbs such as *lician*, Jespersen (1927: 208f.) described this development as a result of two factors, viz. the loss of morphological distinctions, which led to structural ambiguity, and the rigidification of word order, which necessitated a reanalysis of certain constituents. The stages he envisaged for this process are given in (49):

- (49) a. **þam cynge licodon** peran
 the king(DAT-SG) pleased(3-PL) pears(NOM-PL)
- b. the king likeden pears
- c. the king liked pears
- d. he liked pears

Stage a. is a typical example of *lician* in a type-ii construction with a preverbal experiencer, which is analysed as the object of an OVS structure. This constituent order, according to Jespersen, is due to the “greater interest taken in persons than in things” (1927: 208), by which he apparently refers to a greater potential of animate constituents to be topical, a notion that is reminiscent of Elmer’s animacy target for Old English. The second nominal term *peran* ‘pears’ is identified as subject by its nominative case marking and its control of verb agreement on the plural verb form *licodon*, while the experiencer *þam cynge* ‘the king’ is marked as object by dative case. At stage b., this no longer applies to the experiencer, since distinctive case marking is mostly lost from nominal declensions, but the subject-object relation remains overt due to distinctive verbal inflections. At stage c., these are lost as well, and the clause structure becomes ambiguous with regard to the distribution of grammatical relations. As a result of the prevalence of SVO constituent order, the experiencer was then reinterpreted as subject, which is manifested in the use of the subject form *he* at the final stage d. At the same time, a reversal of the verbal meaning apparently took place, since the SVO interpretation requires a receptive glossing of the verb *lician* as ‘like’, while the earlier interpretation presupposes a causative glossing of the verb as ‘please’.

Jespersen’s account was taken up and cast into theoretical terms by Lightfoot in his *Principles of Diachronic Syntax* (1979), which provides a discussion of various historical changes including the rise of a modal category in English and the loss of impersonal constructions. Throughout this treatment, Lightfoot repeatedly emphasises the notion that a

reanalysis of oblique experiencers and similar cataclysmic changes took place as an inevitable consequence of the synchronic structure of language, which was modelled in the general framework of the Extended Standard Theory of generative grammar that was current at the time. The constraining mechanism invoked by Lightfoot (1979: 231f.) in this context is the Transparency Principle, which states that transformational processes by which surface structures are derived from deep structures should always take the simplest possible form, i.e. be as transparent as possible (for a useful discussion cf. McMahon 1994: 116). The actual changes then take place during the process of language acquisition. Language learners construe their mental grammars based on the input that is available in their triggering experience and attempt to construct optimal grammars from this input, which may result in different deep structures compared to those of the original grammars. A surface structure such as (49c), in which the experiencer argument *the king* is not overtly marked as object, will be reanalysed as SVO under the assumption that this is the basic constituent order, since the alternative derivation from basic OVS would involve a double movement of the subject and object NPs, which would be structurally opaque to the language learner, cf. the illustration in (50):



The working definition of reanalysis as a change in underlying structures with no immediate modification to the surface structure can be traced to Langacker (1977: 58), although the scope of this mechanism is discussed in a wider context by him and includes the resegmentation of morpheme boundaries besides the reformulation of more abstract elements of syntax like constituent order (cf. Langacker 1977: 64–92, who also distinguishes several subtypes of these two basic mechanisms). Both types of reanalysis appear to presume some kind of structural ambiguity for which alternative interpretations based on existing models or potentially innovative interpretations are possible. The function of the Transparency Principle is thus to reduce the opacity of the derivation, and while the precise moment at which this occurs is not predictable according Lightfoot, he does assume an absolute limit to the tolerable amount of opacity, which he proposes to establish by inference from observed historical changes (cf. Lightfoot 1979: 121–125). Given the limited number of languages with longstanding historical

records, a fact that Lightfoot himself notes in his preliminary remarks (Lightfoot 1979: 7), the hopes of restricting the Transparency Principle as an explanatory device are not very high, however, and, as it stands, it is neither very well defined nor falsifiable. Whether or not a syntactic structure (or category membership in the case of modal auxiliaries) is opaque enough to require reanalysis remains a matter of debate or needs to be established circularly from cases in which reanalysis is argued to have taken place.

Lightfoot modified his account several times, initially assuming a similar but more concrete syntactic constraint on the development of impersonal constructions, which he called the Trace Erasure Principle (cf. Lightfoot 1981: 225f.). According to this account, the derivation of a surface structure by two opposite moving processes like the ones illustrated in (50) is not only opaque to the language learner but literally impermissible, since NP subjects are considered to require a visible trace such as *it* or *there* after movement. Since this trace would be obliterated by movement of another NP into subject position, movement is barred and reanalysis of the surface structure as SVO becomes necessary. The Trace Erasure Principle is thus not only a descriptive device but the actual cause for reanalysis, which becomes operational as soon as SVO constituent order is established as the base-generated order, a process dated to the 12th century by Canale 1978 (apud Lightfoot 1981: 234). Lightfoot also assumes that the reanalysis of oblique experiencers took place before the final stages of surface ambiguity, adopting a subject analysis of oblique experiencers for Middle English based on sporadic instances of verb agreement with oblique constituents and the double-object construction, which were exemplified and discussed in section 2.1.1.1.4.4 above. A restrictive theory of synchronic grammar is thus considered to account for observable diachronic changes, which would render a dedicated theory of diachronic changes based on historical principles redundant (cf. Lightfoot 1981: 209).

2.2.1.1.2. Criticism

2.2.1.1.2.1. Empirical Basis

Various points of criticism have been raised against reanalysis theory, the most prominent one being its perceived lack of empirical evidence (cf. e.g. Fischer/van der Leek 1983: 337). Indeed, Jespersen's didactic example of the hypothesised process is constructed in a way that allows for a maximum of morphosyntactic distinctions at stage a. and a maximum of ambiguity at stage c. The second constituent *peran* 'pears' is identifiable as the controller of verb agreement by

virtue of its plural number and thus provides additional evidence of the distribution of grammatical relations according to most theories, while no such contrastive evidence would be available with two third-person singular constituents. In addition, the past tense of the verb creates ambiguity at the later stage c., while person marking in the present tense would be distinctive (3. sg. *-s / -th* vs. 3. pl. *-0 / -n*). The survey of complementation patterns of impersonal verbs in section 2.1.3 has also illustrated that type-ii constructions like the one in (49) were not the only type of construction available to experiencer verbs in Old English, and the discussion of the different definitional approaches has made it clear that impersonal constructions, according to the narrower definition as subjectless, would require a nominal stimulus argument in the genitive case. Strangely, the latter point is entirely absent from any discussion of reanalysis, despite the fact that genitive case marking or the alternative realisation by a PP constituent generally remain available throughout the history of English. Early exceptions to the distinction of genitive case marking are the singular forms of strong feminine nouns and all weak declensions in Old English, but these merged into a basic noun declension in Middle English with a generalised genitive singular ending *-(e)s*. Not even van der Gaaf mentions this point, even though his goal of providing empirical evidence for the theory of reanalysis required a more detailed survey of impersonal verbs, which includes verbs like *hreowan* ‘to rue’ that are attested with a genitive complement. All of his citations of this verb contain either a nominative constituent, pronominal *hit*, or a clausal complement (cf. van der Gaaf 1904: 5f.). The lack of citations of genitive complements might be an oversight or a deliberate simplification, but the ambiguity of clausal constituents with regard to case marking is most likely a contributing factor in the typical neglect of genitive constituents in the context of reanalysis theory (but cf. the brief discussion in Elmer 1981: 75, 77). At the same time, the morphological ambiguity of clausal constituents suggests that more verbs than the typical type-ii verb *lician* could have been affected by reanalysis, even if they otherwise entered into type i, since the distinction between the type i and ii was neutralised with clausal complements. In its stated form with a second nominal complement, however, reanalysis theory only attempts to account for the development of the experiencer argument, not for that of the stimulus argument.

2.2.1.1.2.2. Ambiguity and Position of the Experiencer

The emergence of structural ambiguity as a result of non-distinctive case marking and the experiencer’s propensity towards clause-initial or at least preverbal position have both been called into question (cf. Allen 1986a: 378–380, McCawley 1976: 198). This presents a potential

challenge to reanalysis theory, since structural ambiguity and the reassignment of grammatical relations according to basic SVO constituent order are its two fundamental tenets, even though structural ambiguity plays a less crucial role in Lightfoot's account. As is well known, the distinction between nominative and oblique case (or "subject" and "object" case) is maintained in the domain of English personal and anaphoric pronouns (cf. ME *I* vs. *me*, *thou* vs. *thee*, *we* vs. *us*, *ye* vs. *you*, *he* vs. *him*, *she* vs. *hir*, *they* vs. *hem* and their dialectal variants). This circumstance was also acknowledged by Lightfoot (1979: 232), who simply claimed, in his earlier account, that "[s]uch was the force of the Transparency Principle, to require a canonical SVO analysis and the simplest transformational derivation, that even *him liked the pears* was analysed as SVO despite the form of the pronoun". While this statement may seem little too optimistic about the apodeictic nature of its claim, it does highlight the difficulty of establishing the degree of morphological syncretism that would result in sufficient structural ambiguity for an alternative or innovative syntactic analysis to (necessarily) take place. This basic problem severely limits the explanatory force of the Transparency Principle, but it applies equally to other lines of argument, whether they be in favour of or against the role of structural ambiguity as a trigger of reanalysis. It nevertheless seems reasonable to assume that the frequency of ambiguous contexts, measured in types or, more likely, tokens, is an important factor in any effect that structural ambiguity may have on diachronic change. For experiencer verbs in Early Middle English, such contexts arise with nominal third-person experiencers in the singular and plural if these are not disambiguated by verb agreement, but generally not with pronominal first- or second-person experiencers. In narrative texts, which constitute the dominant source of Middle English, the third person may, in fact, be relatively frequent, but this is less representative of the spoken language, which presumably exhibits a much higher frequency of first- and second-person constituents. Since spoken language arguably constitutes the typical starting point for most types of language change, third-person experiencers like the one in Jespersen's example formed after Old English narrative texts may be less relevant to the development. In a similar vein, McCawley (1976: 198) argues that first-person experiencers are expected to be more frequent with most impersonal verbs, since they express "inherently personal subjective experiences".

Even in narrative texts, however, which provided the inspiration for early reanalysis theory, examples of third-person experiencers in ambiguous contexts appear to be not very frequent. A count conducted by Allen (1986a: 378f.) found that only 10 out of 110 instances of *lician* with two nominal complements did not contain at least one pronominal realisation that disambiguated the grammatical relations in the prose section of Healey/Venezky's

Concordance to Old English, not counting 5 additional instances with PP experiencers. The equivalent figures for her collection of EME texts and Tatlock/Kennedy's *Concordance to the Complete Works of Geoffrey Chaucer* are 4 out of 27 instances and 1 out of 32 instances respectively. These numbers suggest that patterns like *the king liked pears* in the constructed example in (49) above were not very typical during any of the investigated periods of English. One could, of course, still argue that a less frequent pattern may be sufficient in triggering reanalysis. Since Allen's investigation regards instances with a distinctively marked stimulus arguments as unambiguous cases, the incidence of morphologically ambiguous experiencer arguments is actually higher than the above figures suggest, and these instances would not be susceptible to the same kind of disambiguation in type-i constructions with a genitive complement. More importantly, however, the count excludes instances with clausal complements, which are equally not suited to the disambiguation of nominal experiencers, and whose frequency, in certain expressions, potentially results in a much higher incidence of morphosyntactically ambiguous experiencers. More specific numbers of ambiguous constructions of individual verbs will be discussed in section 4 of the present study.

The same data restriction affects Allen's second point about the experiencer's position, which she considers to have been in line with an object interpretation throughout the history of English (cf. Allen 1986a: 279). Her point seems surprising at first, since she argues in favour of a subject analysis of oblique experiencers based on their frequent preverbal position elsewhere (1986b: 469). The discussion of Elmer's syntactic classification of impersonal verbs has already illustrated, however, that multiple constituent orders were available in type-ii constructions with two nominal complements, in which either the stimulus argument (Elmer's type Ia) or the experiencer argument (Elmer's type Ib) could precede the other. Allen's count of OVS and OSV order with *lician* and its EME and ME successors suggests that these were always less frequent than the alternative SVO or SOV patterns, the latter of which was generally favoured in Old English with pronominal objects and at least optionally present in subordinate clauses with nominal objects. The exact figures for the former constituent orders are 50 out of 110 instances (45%) in Old English, 7 out of 26 instances in Early Middle English (27%) and 6 out of 30 (20%) instances in Chaucer (the total numbers differ slightly compared to the figures above due to additional restrictions on the dataset). The situation is altered drastically, however, when instances with clausal complements are considered, which always occur in final position in the case of finite subordinate clauses, and which typically occur in final position in the case of non-finite subordinate clauses as well. Overall, Allen's investigation does reveal that Jespersen's hypothetical example of a type-ii construction with an ambiguous experiencer

argument in preverbal position is not equally representative of all experiencer verbs and the different constructions available to them, but if instances with clausal complements are included, the potential for structural ambiguity and preverbal position of the experiencer are actually increased.

Empirical evidence of ambiguous constructions is cited by van der Gaaf (1904: 25–36) and Visser (1963: 30f.), whose examples of type-ii constructions often contain a proper noun accompanied by a clausal complement as in (51):

(51) **Juno** list nat at the feste be

‘Juno did not desire to be at the festivity’ (Chaucer, *L. G. W.* 2249)

Such proper nouns are probably exceptional in combining a lack of inflection with a high potential to occur as experiencers, which results in their noticeably high frequency in narrative contexts. Yet there are further constructions which van der Gaaf (1904: 25–36) and Jespersen (1894: 219–223) consider influential in triggering reanalysis. The first of these is the realisation of the experiencer by an indeclinable pronoun or particle, mostly prominently by relative *that* and indefinite *whoso*, but also relative and interrogative *who*, for which the distinction between a subject form *who* and an object form *whom* was already starting to become blurred in Middle English (cf. Jespersen 1949: 241–244). To these, the relative particle *the* (OE *þe*) can be added, which equally does not exhibit case marking. An example of an experiencer argument realised by one of these indeclinable items is given in (52):

(52) nas never noon **that** luste bet to singe

‘there never was anyone more inclined to singing’ (Chaucer, *Cant. T.* G 1344)

Other contexts of structural ambiguity are the accusative-with-infinitive construction (“subject-to-object raising”) and victimhood to coreferential deletion. Examples of these contexts were given in (17) and in (13) and (14) respectively, to which (53) and (54) below can be added:

(53) than preye I **thee** to rewe up-on my pyne

‘then ask I you to take pity upon my strife’ (Chaucer, *Cant. T.* A 2382)

(54) god alloweþ 3our lif & likeþ 3our dedes

‘God_i approves of your life and _____i likes your deeds’ (*Alex. & Dind.* 212)

While such instances of complex sentences may not have been the most frequent type of construction across all registers, they present additional contexts in which structural ambiguity may have arisen. Also the fact that not only nominal but also some pronominal items did not consistently distinguish between subject and object case increases the potential role of structural ambiguity as a basis for reanalysis. In combination with the experiencer's occurrence in preverbal position, especially in connection with clausal complements, the basic requirements of reanalysis theory do seem to be met, although their impact on the development of oblique-experiencer verbs and the question of whether an absolute majority of ambiguous contexts is required for reanalysis to take place remain disputed issues.

2.2.1.1.2.3. Instantaneity of Change

Further problems appear to be attached to Lightfoot's account in particular, since the restrictive nature of the Trace Erasure Principle and the attribution of diachronic change to an abductive process during language acquisition require that reanalysis took place instantaneously once the precondition of base-generated SVO constituent order had been established. The expectation that a reanalysis of underlying structures is able to account for a range of simultaneous changes in their surface manifestations is one of the factors that makes reanalysis an attractive explanatory device in Lightfoot's view, since the cataclysmic nature of such changes and their simultaneous occurrence would otherwise be difficult to explain. Neither of these two predictions is met by the data, however. Most impersonal verbs continue to be attested with oblique experiencers until the end of 15th century or, in some cases, beyond. If the establishment of base-generated SVO is dated to the 12th century, this creates a conspicuous gap between the supposed trigger of reanalysis and its ultimate effects, quite contrary to Lightfoot's statement that reanalysis occurred "hard on the heels" of the constituent-order change (cf. Lightfoot 1981: 225 and the critique by Fischer/van der Leek 1983: 342). This problem was, of course, also noticed by Lightfoot himself but apparently not considered important enough to interfere with his theory (cf. Lightfoot 1979: 232¹). Regarding the second point, it is quite clear that the changes to impersonal verbs did not affect all of the relevant items simultaneously and in a cataclysmic fashion. This emerges from Elmer's charts of ME *rue*-verbs, for example, which show that oblique experiencers continue to occur in preverbal position in conjunction with clausal complements until the 16th century for *reuen* and until the 15th century for *shamen* and *forthinken*, while *greven* requires expletive *it* or a nominative experiencer after the 14th century (cf. Elmer 1981: 88 and 104). With nominal complements, all of the above verbs continue to

be optionally used with an oblique experiencer until the 16th century except *shamen*, which only does so until the 15th century in Elmer's collection of examples. Further differences arise from the variable realisation of the stimulus argument by a NP or a PP. The development of these verbs is thus far from uniform, even within a single semantic class.

2.2.1.1.2.4. Productivity

Another problem with the notion that preverbal oblique experiencers were reanalysed as subjects as a result of their object analysis becoming structurally impossible lies in the fact that constructions with oblique experiencers not only continued to exist but also attracted new members to the class of impersonal verbs during Middle English, a fact which seems difficult to account for if reanalysis, indeed, constituted a structurally required replacement. As already mentioned in the introduction (cf. section 1.2), this process affected loan words from Old Norse and Old French, which represent the most influential donor languages for Early English besides Latin. The verb ME *dremen* 'to dream', for example, which replaced OE *mætan* 'id.', most probably constitutes a loan of ON *dreyma* 'id.' rather than a continuation of OE *dreman* 'to rejoice, play an instrument', given the difference in meaning (the latter is continued by ME *dremen* 'to resound'). The impersonal use of this loan word was, of course, favoured by the fact that both ON *dreyma* and OE *mætan* were already constructed with an oblique (accusative) experiencer. Similarly, ON *gegna* 'to pertain, befit', which is the source of ME *geynen* 'to avail, be profitable', was used with a dative experiencer (or rather a dative beneficiary) already in Old Norse and continued this use in Middle English. A number of loan words from Old French, however, were originally used in reflexive constructions both in Old French and in Middle English before their attestation in impersonal constructions. An example is ME *remembren*, derived from OF *remembrer*. Van der Gaaf (1904: 144f.) notes the impersonal use of this verb as one of the peculiarities of Chaucer's language besides the more common reflexive and transitive constructions. There are other verbs like ME *repenten* 'to feel regret', however, whose innovative construction with an oblique experiencer compared to its source OF *repentir* 'id.' became a more general phenomenon during the 15th century in conjunction with expletive *it*, and which continued besides reflexive and transitive use into the Elizabethan era (cf. van der Gaaf 1904: 152f.). The co-existence of alternative constructions involving oblique experiencers on the one hand and constructions with a reflexive pronoun on the other indicates another dimension of complexity in the syntactic complementation of impersonal verbs, which will not be explored in detail in the present study (for a brief survey of verbs exhibiting impersonal and

reflexive constructions in Old English and Middle English cf. Möhlig-Falke 2012: 187–192 and 222f. respectively). From a functional point of view, reflexive constructions seem to be related to constructions with oblique experiencers in that they do not encode a prototypical agent like transitive constructions. Under the assumption that the syntax of Middle English structurally required preverbal subjects that were typically marked by nominative case, the introduction of oblique experiencers besides functionally related constructions that fulfilled these criteria seems rather surprising.

Another way in which Middle English acquired additional impersonal verbs was by word formation or the reinterpretation of existing verbs. The verbal compound ME *forthinken* ‘to regret, repent’, for example, replaced the compound *ofthinken* ‘id.’ during the 13th century, but it continued the impersonal construction of the latter, which was inherited from OE *ofþyncan* ‘displease’ (cf. van der Gaaf 1904: 13–15). Reinterpretations are particularly well attested for verbs expressing modal meanings of obligation and necessity. Sporadic impersonal occurrences are noted by van der Gaaf for ME *moten* ‘to be allowed’, which occurs with an oblique experiencer in its secondary sense ‘to be compelled’ (cf. van der Gaaf 1904: 143f.). Similarly, ME *ouen* ‘to possess’ is used impersonally in its figurative sense ‘to be suitable, proper’, and the same is true of ME *thurven* ‘to need’. Both of these verbs occur in impersonal constructions besides personal constructions until around the end of the 15th century (cf. van der Gaaf 1904: 146–148, 154). Finally, impersonal use of the personal verb OE *behofian* ‘to need’ is adopted with the obligational sense of ME *bihoven* ‘to be needed, proper’, whose semantic development has been described as a change from “internal” to “external necessity” (cf. Loureiro-Porto 2010: 695, on its development from personal to impersonal constructions cf. also Allen 1997). A motivation for this extension of impersonal constructions to verbs of obligation in particular seems to be the analogy of semantically related impersonal verbs such as ME *semen* ‘to be fitting’ and *neden* ‘to be necessary’, a point also noted by van der Gaaf (1904: 144, 146). While verbs of obligation thus constitute a particularly prominent semantic class in this development, the cases of *remembren* and *repenten* show that verbs of cognition and verbs of emotion could equally be affected. Innovative impersonal use across a range of semantic classes seems difficult to account for under Lightfoot’s assumption that preverbal constituents were obligatorily analysed as subjects after base-generated SVO had been established, since it suggests that oblique subjects were not only tolerated but also actively introduced besides syntactically regular nominative subjects despite the opacity of their grammatical status. The restrictiveness of Lightfoot’s diachronic account, which was intended as its primary merit, thus becomes a major flaw when compared against the empirical evidence.

In addition, the observable presence of variation between personal, impersonal and reflexive constructions in the works of a single author such as Chaucer argues against the notion of an instantaneous low-level change during the abductive process of language acquisition, since it does not account for the observed intra-individual variation. It seems, then, that the development from oblique experiencers to nominative experiencers cannot be accounted for in terms of a cataclysmic reanalysis, at least not without additional assumptions.

2.2.1.1.2.5. Lexical Diversity

A final criticism of reanalysis theory lies in the fact that different impersonal verbs take different pathways in their development. Similar to its input, the outcome of the diachronic change is more complex than the didactic example of reanalysis presented by Jespersen suggests. Apart from the personalisation of impersonal constructions by the introduction and increased use of nominative experiencers, two major developments are commonly distinguished, viz. the obsolescence of impersonal verbs and their syntactic adaptation in the form of *it*-constructions. The former development seems less problematic for reanalysis theory, since it does not affect the structural argument, although it does seem to imply that impersonal constructions were disfavoured regardless of structural necessity. It is not easy to distinguish between cause and effect in this scenario, however, since an overall decline of impersonal constructions might lead to the obsolescence of individual verbs, but the obsolescence of individual verbs may equally be seen as a contributing factor to the general decline. The question is thus whether impersonal verbs were affected more severely by lexical obsolescence in comparison to other verbs, which could indicate that obsolescence affected impersonal verbs in particular. This does not seem to be the case with regard to verbal compounds like OE *forsceamian* ‘to make ashamed’, OE *gebyrian* ‘to happen’ and OE *gedafenian* ‘to benefit’ (cf. Visser 1963: 29), which were frequently lost from English regardless of their status as personal or impersonal verbs. Simple impersonal verbs that became obsolete include OE *preotan* ‘to weary’ as well as simple verbs besides compounds such as OE *limpan* ‘to happen’ besides OE *gelimpan* ‘id.’ and OE *tweogan* ‘to cause doubt’ besides OE *getweogan* ‘id.’. The fact that impersonal constructions were productively adopted for a range of other verbs during the ME period seems to suggest, on the other hand, that lexical obsolescence itself was not a causative factor in the demise of impersonal verbs, since the subset of impersonal verbs both lost and gained individual items during its development. The fact that some impersonal verbs were lost early while others were only

acquired later certainly makes it more difficult to trace the overall development, however, since the data of individual items can be limited to particular periods.

The continuation of impersonal verbs in the form of *it*-constructions like PDE *it seems to me* and *it behoves us* presents a more interesting case, since these constructions satisfy the requirement of unambiguous SVO constituent order while maintaining the oblique coding of the experiencer argument, either in the form of a PP or in the form of a constituent marked by oblique case. A reanalysis of preverbal experiencer arguments as nominative subjects was thus not a necessary consequence of the establishment of base-generated SVO (cf. also McCawley 1976: 192f., von See Franz-Montag 1984: 529). Lightfoot was, apparently, not concerned by such lexically diverse developments, stating that diachronic predictions about individual lexical items were “impossibly ambitious” (1979: 232¹). The viability of an alternative construction besides personalisation nevertheless leaves the assumption of a structurally required reanalysis open to criticism. There are certain restrictions to the distribution of expletive *hit*, however, which partly account for the different developments. First of all, the argument structure of a given verb plays an obvious role for zero-argument verbs like *sniwan* ‘to snow’, which typically occur with an expletive already in Old English, while one-argument verbs with an oblique experiencer like *hyngrian* ‘to hunger’ do not. Second, the syntactic realisation of the stimulus argument affected the distribution in that *hit* was much more likely to occur with clausal complements than with nominal complements, and even in the former case, it was rare if an experiencer argument was expressed at the same time (recall that Allen’s count of *gelimpan* ‘to happen’ included no instance of *hit* besides an experiencer argument in main clauses and only one instance in subordinate clauses, and also Elmer’s example of the variant of type S with an experiencer and an expletive actually contained demonstrative *þæt*). Two-argument experiencer verbs were thus initially less likely to acquire the *it*-construction, particularly if the second argument was realised by a nominal complement.

The scope of expletives increased during Middle English and extended to constructions with an experiencer, which is often realised by a PP in such cases. The examples collected by van der Gaaf (1904: 36–39) include instances of *semen*, which continues to be used with an expletive pronoun in Present-Day English, but also instances of *liken*, which does not continue this construction, cf. (55) and (56) below:

(55) *hit semis vnto vs þai are feler of folke*

‘it seems to us they are more numerous with regard to their people’ (*Destr. of T.* 4868)

(56) it lyketh to your fader and to me that I yow wedde

‘it pleases your father and me that I marry you’ (Chaucer, *Cant. T.* E 345)

These examples differ from the case of OE *gelimpan* in that the preverbal position is filled by an expletive despite the availability of an experiencer argument for that position. As such, they seem to present an ideal compromise between impersonal constructions and the assumed constraint on constituent order, since the oblique status of the experiencer is indicated by its realisation as a PP, while the expletive fulfils the requirement of SVO with a nominative subject. The question is, then, why constructions like (56) were not preserved, even though they were not under pressure from the base order (cf. also von Steffens 1984: 529). A possible reason lies in the fact that ME *liken* was also used in type-ii constructions with two nominal complements, which did not favour the presence of an expletive pronoun, while ME *semen* presupposed a propositional statement in the form of a finite or non-finite subordinate clause, which continued the typical construction with a raised (and potentially empty) subject. There were, thus, contexts entered by *liken*, but not by *semen*, in which the adaptation to basic SVO constituent order required a nominative experiencer, and these contexts may have been instrumental in the abandonment of the alternative construction with expletive *it* for *liken*.

2.2.1.1.3. Summary

The restrictive account of the diachronic development of impersonal verbs proposed by Lightfoot encounters significant problems when compared to the empirical evidence. Jespersen’s didactic example of the assumed mechanism of reanalysis certainly does not represent the entire scope of impersonal constructions, but while type-ii constructions with two nominal complements may be more variable in their constituent order and possess a higher potential for structural disambiguation by pronominal experiencer or pronominal stimulus arguments, the conditions for surface reanalysis seem to have been rather more favourable than less favourable in type-i constructions with a genitive complement and generally with clausal complements. In any case, the amount of structural ambiguity that is necessary or sufficient for triggering reanalysis remains a difficult issue, both theoretically and empirically. Even more problematic is the fact that the assumed consequences of reanalysis are neither as immediate nor as rapid as the assumed initiation of the abductive process during language acquisition by the establishment of base-generated SVO would seem to predict according to the proposed Trace Erasure Principle. Impersonal verbs continue to be productive during the ME period by

adaptation of foreign loan words as well as by reinterpretation of indigenous verbs, most prominently verbs of obligation. Also, not all experiencer verbs are subject to reanalysis, but some simply become obsolete, while others develop a stable construction with expletive *it*, which suggests that reanalysis does not account equally for all lexically diverse developments. Whether the ambition of a diachronic account should include such lexical details may be a matter of preference, but it needs to be acknowledged, nevertheless, that a seemingly parsimonious account falls short of a comprehensive representation of the object of interest if it neglects such details. A less restrictive account, while theoretically less attractive, may thus prove to be more adequate overall. For the development of impersonal verbs, at least, it seems that the apparent gradualness and lexical diversity of the change are important aspects, which need to be accounted for.

2.2.1.2. Lexical Functional Grammar

2.2.1.2.1. Outline

Allen's account of impersonal verbs in the framework of Lexical Functional Grammar (1995) deserves special attention because it is based on a much more detailed investigation of the empirical evidence than the initial proposal of reanalysis and, in many respects, presents an explicit alternative to the latter theory. A fundamental characteristic of the framework adopted by her is the abandonment of syntactic transformations (although morpho-lexical operations exist) in favour of a set of general principles that combine with lexical specifications, which are intended to account for observable patterns in the constituent structure. The general principles include a hierarchy of semantic (or "thematic") roles, which partly determines their mapping onto grammatical relations, as well as a classification of grammatical relations in terms of two syntactic features as $\pm r$ (i.e. restricted or unrestricted in their lexical mapping) and $\pm o$ (i.e. object or non-object). Similarly, semantic roles are assumed to be associated with either intrinsic or default role classifications, which restrict their mapping onto grammatical relations. With regard to case marking, a general distinction is made between structural assignment, which takes place according to the general principles, and lexical assignment, which results in idiosyncratic patterns. Allen specifically assumes structural case marking to take place according to a case hierarchy, which is intended to capture the assumed object status of postverbal nominative constituents in Icelandic. These constituents receive nominative case as the first available structural case after its assignment to the experiencer argument in the relevant constructions has been overridden by the lexical assignment of dative case (cf. Allen 1995:

125f.). A similar constraint is assumed for experiencer arguments in Old English according to the thematic hierarchy, which predicts that the higher-ranking experiencer argument can only be assigned lexical dative in the presence of a nominative theme but not structural accusative, since it would have to be assigned nominative first (cf. Allen 1995: 129). Syntactic generalisations are thus formulated in terms of a relatively limited set of rule combinations, while all other facts of the observable constituent structure must be specified lexically.

In her investigation of the case patterns of OE experiencer verbs, Allen largely follows Elmer's terminological system and his basic distinction between constructions with two nominal complements and constructions with a nominal and a clausal complement (cf. section 2.1.3.2 above and Allen 1995: 68–88). The former are collectively referred to by Allen as “2NP types”, which include the impersonal type N and its two variants type I and type II, while the latter are referred to as “PROP types”, which include the impersonal type S and its variants with a nominative experiencer (“personal type”) or an expletive pronoun (“type *hit*” and “type DEM”). In addition, Allen distinguishes a construction with an understood propositional argument that is grammatically unexpressed, which she refers to as “NO PROP” construction (cf. Allen 1995: 276 *et passim*). For the sake of consistency, the more straightforward numbering of the constructional types as i–iii will be used here, but since the focus of Allen's discussion is on the assignment of case to NP arguments, and since a distinction of type-i and type-ii constructions is not meaningful for clausal complements, these terms should be understood to refer to the 2NP types primarily. Attestations of other potential case patterns like the combination of a dative experiencer and an accusative stimulus argument (cf. e.g. *him gelicade hire þeawas* in (18) above) are dismissed by Allen on philological grounds, since alternative explanations are mostly available for these cases (cf. Allen 1995: 74–79). The remaining patterns are accounted for by a combination of the instruments available in the general theoretical framework and any additional specifications that the empirical evidence requires. For example, Allen assumes that accusative case is assigned lexically to experiencer subjects and thus restricted to -o arguments, since this allows for the preservation of the case hierarchy, which would require the assignment of nominative case if the assignment were structural, while also accounting for the fact that accusative case is generally not preserved under passivisation, which would be expected if accusative case could be assigned lexically to objects (cf. Allen 1995: 138f.). The impersonal type i (Allen's and Elmer's type N) can thus be represented by a lexical frame in which the experiencer subject, which is marked as -r on the basis of its syntactic subject properties, can receive either lexical accusative or dative case, while the second argument, the stimulus (or “Theme” in Allen's treatment), receives either

genitive case as a result of being marked +o, or is realised by a PP as a result of being marked +r, both of which are structurally related with the respective syntactic features (cf. Allen 1995: 136–140). The variable realisation of the stimulus argument appears to be a weak point in this treatment, especially since the classification of the stimulus according to mutually exclusive syntactic features is delegated to the speaker's choice, although the mental reality of such a process seems questionable, but since the variation between genitive and PP complements is not very well accounted for by reanalysis theory either, this weakness is at least not specific to Allen's account. A more important point is the variation between nominative and oblique experiencer arguments or between type-i and type-iii constructions. Allen, similar to Fischer/van der Leek (1983), proposes to explain this variation with the assumption that lexical case assignment to the experiencer argument was optional, and that experiencers could also receive structural nominative case by virtue of representing the highest-ranking semantic role in their lexical frame. There are two important differences compared to Fischer/van der Leek's account, however. First, no general optionality of lexical case marking is assumed, which would involve a conflation of all three constructional types in a single lexical frame. Second, the optionality of lexical case marking of the experiencer argument is assumed to be lexically restricted to specific items, while other lexical items are assumed to receive lexical case marking obligatorily (e.g. *lystan* 'to cause pleasure').

Type-ii constructions (Allen's and Elmer's type I) receive a different treatment, which is intended to capture the evidence from deletion under coordination that preposed oblique constituents could be syntactic subjects in Old English. Based on her investigation of a range of additional properties, which include pronominal realisation, animacy, and the distribution of the expletive, Allen analyses instances with a preceding experiencer argument (Elmer's type Ib) as experiencer subjects, while instances with a preceding stimulus argument (Elmer's type Ia) are analysed as stimulus subjects. In order to account for this variation theoretically, Allen assumes that the experiencer is marked -r by default and that the stimulus is marked -r intrinsically, which results in the availability of both of these arguments for the role of either subject or object. The experiencer always receives dative case lexically, while the stimulus argument is structurally marked by nominative case, either as subject or, when the experiencer is subject, as the only argument that receives structural case marking, similar to the assignment of nominative case to postverbal constituents in Icelandic discussed above. Variation between type-i constructions and type-ii constructions is accounted for by two separate lexical entries, mainly on the grounds that the assumption of general optionality of lexical case marking would also permit transitive patterns, in which both arguments receive structural case, while such

patterns do not seem to be attested for *ofhreowan* ‘to cause pity’, which otherwise exhibits all three constructional types. Instead, the lexical case marking of the experiencer argument of *ofhreowan* is assumed to be obligatory in type-ii constructions with a nominative stimulus argument, but it is regarded as optional in constructions with a stimulus argument that is syntactically marked and consequently realised by a genitive or PP complement, which produces type-i and type-iii constructions. Optionality of lexical case marking is thus treated as a lexical rather than a structural feature, which permits the representation of a great deal of the attested lexical variation, including lexical items like *hreowan* ‘to rue’, which potentially requires such optionality for the experiencer even in absence of syntactical marking of the stimulus argument in order to account for variation between type-ii constructions and transitive constructions if the (relatively scarce) evidence of the latter type of construction is genuine (cf. Allen 1995: 150–154).

The lexical restriction of the optionality of lexical case marking and the conditions imposed upon it imply that type-ii verbs will typically exhibit only a single constructional pattern apart from any variation between nominal and clausal complements, although the subject status of the oblique experiencer may vary according to its position relative to the stimulus argument. A typical example of such a type-ii verb with variable assignment of grammatical relations is OE *lician* ‘to please’. In addition, Allen assumes the existence of so-called “Dative Object verbs”, represented by OE *cweman* ‘to please’, whose experiencer argument is consistently assigned the role of object. The analysis is motivated both by the failure of such verbs to develop nominative experiencers diachronically and by a set of synchronic properties that are similar to those adduced in favour of the alternative object analysis of experiencer arguments in type ii, including the postposed position of the experiencer relative to the second argument and its apparent lack of control of deletion under coordination besides a more frequent realisation of the second argument by a pronoun and its reference to animate participants (cf. Allen 1995: 144–149). It should also be mentioned, however, that the model verb *cweman* is one of the verbs that becomes largely obsolete in the EME period and thus does not exhibit the full development which it may otherwise have taken. In addition, it is also not particularly frequently in Allen’s collection of OE texts, although the somewhat more frequent attestations of the compound verb *gecweman* ‘to please’ appear to point in a similar direction. The different behaviour of the different types of verbs is described as part of a diachronic scenario, in which the option of assigning the role of subject to the experiencer argument was acquired at a later stage for type-ii verbs than for type-i verbs, since the latter are arguably more progressive in Old English in allowing optional structural case marking of the

experiencer subject, while lexical case marking of the experiencer argument is assumed to be still largely obligatory for type-ii verbs (cf. Allen 1995: 153). The difference between type-ii verbs and Dative Object verbs, on the other hand, is explained by a difference in the semantic role of the second argument, which will be referred to as stimulus in the remainder of this investigation, but for which Allen makes a distinction between theme and cause argument. This distinction is reminiscent of Anderson's analysis (1988) of the second argument as either ablative or ergative, although he adopts the latter analysis for the nominative constituent of type ii in contrast with types i and iii. For Allen, the identification of the second argument of Dative Object verbs as cause is intended to capture the fact that the oblique experiencer of such verbs does not appear to exhibit syntactic subject properties. This can be explained with reference to the assumed hierarchy of semantic roles in that only the experiencer argument of type-ii verbs may assume the role of subject by default, given that it outranks the theme argument of these verbs but not the cause argument of Dative Object verbs (cf. Allen 1995: 147).

Allen is well aware that her distinction of different semantic roles for the second argument as a way to account for the apparent failure of the experiencer argument of verbs like *cweman* to exhibit the same subject properties assumed for preposed oblique constituents in type-i constructions and for preposed nominative constituents in type ii constructions runs the risk of circularity. The distinction is therefore grounded in a larger investigation of apparently independent features, which are intended to substantiate the semantic motivation of the contrast. Apart from the statistical difference in the relative position of the theme and the experiencer argument on the one hand and the cause and the experiencer argument on the other, the cause argument is assumed to exhibit a more volitional interpretation of what Allen refers to by the more general term "Target of Emotion", which would motivate the consistency of its subject analysis in comparison to the optional subject analysis of the theme argument of type-ii verbs (cf. Allen 1995: 144–149). The primary indication of this volitional interpretation is taken to be the predominant reference of the second argument to animate participants. Out of 8 occurrences of *cweman* and 29 occurrences of *gecweman* in constructions with two nominal arguments, only 1 exhibits a non-human stimulus in Allen's count, in addition to which she notes at least one example of *cweman* in which the experiencer argument is unexpressed. This contrasts with a much higher proportion of non-human stimuli in comparable constructions of *lician* (82 or 79% out of 102 instances according to Allen's data). Upon individual examination, Allen finds that all instances of *(ge)cweman* with two nominal complements are consistent with a volitional interpretation of the second argument, although the deviant example of a 2NP construction with a non-human stimulus argument (*ba word* 'the words') seems to permit this interpretation only

indirectly, and the instance of a non-human stimulus argument without an explicit experiencer (*þes middangeard* ‘this world’) illustrates, according to Allen, that the volitional interpretation did not apply by necessity. Additional indications of a predominantly volitional interpretation are the frequent prefixation of *cweman* by *ge-*, which, according to Lindemann (1970 apud Allen 1995: 146⁵⁵), indicates a directed and hence intentional action, as well as individual occurrences that function as complements of *willan* ‘to want’, which explicitly states the volitional nature of the act of pleasing. Similar instances of complementation of *willan* are attested for *lician*, which means that the distinction between a more volitional cause argument for *(ge)cweman* and a more neutral theme argument for *lician* is based on statistical rather than absolute evidence and allows for some intra-lexical variation. The final point is the categorial realisation of the second argument. The figures presented by Allen, which apparently include one additional instance of *cweman* with two nominal constituents compared to the instances underlying her investigation of animacy, show a relatively clear tendency towards a more frequent pronominal realisation of the second argument of *(ge)cweman* in comparison to both the experiencer argument and in comparison to the respective categorial realisation of the two arguments of *lician*, for which the experiencer argument is more frequently pronominal. Allen considers this criterion of weight “an extremely crude measure” (Allen 1995: 147), whose main function is to represent the respective tendency towards topicality, which is found to be greater for the experiencer argument of *lician* and greater for the stimulus argument of *(ge)cweman*. In this way, the criterion ties in with the proposed criterion of a volitional interpretation based on animacy, since the constructions using *(ge)cweman* can thus be said to be (generally) more about the volitional act of the stimulus than the equivalent constructions of *lician*.

The semantic motivation for the different formal analyses of type-ii verbs represented by *lician* and Dative Object verbs represented by *(ge)cweman* also plays a role in Allen’s diachronic account of the development of impersonal verbs. The principal argument is repeated and elaborated by Allen in a later part of her study (cf. Allen 1995: 326–347), where she extends the account to the apparent absence of ME passive constructions of *liken* in which the experiencer rather than the stimulus argument constitutes the subject of the passive clause (i.e. *he was liked* in the sense of ‘he was pleased’). Such constructions can be referred to as causative passives, since they presuppose a causative interpretation of the verbal semantics in the corresponding active clause rather than a receptive interpretation (cf. the terminological distinction with regard to active clauses by Fischer/van der Leek 1983: 354), and the apparent absence of causative passives in Middle English can be explained theoretically by the assumption that only the experiencer but not the lower-ranking theme argument of the relevant

verbs is suppressed under passivisation (cf. Allen 1995: 335). The semantic basis of this difference compared to Dative Object verbs, which do exhibit suppression of the higher-ranking cause argument in corresponding causative passives, not only motivates the partly subject-like syntactic behaviour of the experiencer argument of type-ii verbs, but is also taken to underly the extension of coding properties in the form of structural nominative case marking to oblique experiencer arguments. A case in point is the OE verb *losian*, for which a semantic shift from ‘to perish’ towards ‘to go astray’ is assumed (cf. Allen 1995: 326–328). This verb is the only one that Allen identifies as undergoing a relatively early and rapid shift from oblique to nominative experiencers, which would thus conform to the proposal of reanalysis theory. In keeping with the semantic hypothesis and in view of the apparent lack of a structural change, she instead connects the development to the fact that the secondary sense of *losian* can imply responsibility of the owner for what is being lost. This responsibility, although obviously not to be equated with volition, is supposedly reflected in the early extension of nominative case to the relevant arguments. Along similar lines, Allen draws a distinction between occurrences of unambiguous nominative experiencers in type-iii constructions of *reuen* in the writings of Chaucer, which she glosses as ‘to take pity on’, and a single instance of *reuen* in a type-i construction with an oblique experiencer, which she glosses as ‘to feel pity for’. The nominative case marking of the former instances is viewed as a clear indication of a “more active sort of Experiencer” (cf. Allen 1995: 309). These instances also illustrate that the distinction of different semantic interpretations, which Allen largely identifies along the lines of different lexical items, also involves some degree of intra-lexical variation. A similar example of this variation is Chaucer’s occasional use of *liken* to signify an apparently intentional act (cf. [w]el lever is me liken yow and deye ‘I would much rather please you, and die’, cited by Allen (1995: 337)). While the principal difference is thus perceived between type-ii verbs with variable subject assignment and a less responsible theme argument on the one hand and Dative Object verbs with consistent subject assignment and a more responsible cause argument on the other, the statistical nature of this difference and the assumption of a gradual process of change necessarily entail the existence of intra-lexical variation.

Two additional contexts in which a semantic explanation for the preverbal placement and the adoption of nominative case marking for experiencer arguments seems to be available are the so-called NO PROP constructions of ME *plesen* ‘to please’, in which a propositional argument is implicit, and, to a lesser degree, PROP constructions of the same verb, in which a propositional argument is explicit (cf. Allen 1995: 338–347). For the former, Allen assumes an additional distinction between occurrences in “variable clauses” introduced by *how*, *as*, *what*

and similar conjunctions on the one hand and occurrences in conditional clauses introduced by *if* or *and* on the other, by which she accounts for much of the variation observed in Shakespearean usage of the verb. In Spevack's *Concordance to Shakespeare*, constructions with expletive *it* and postverbal position of the experiencer are much more common in the latter context, while the former context exhibits much fewer comparable instances and a higher proportion of unambiguous nominative experiencers, although morphologically ambiguous experiencers are more frequent in either context. The semantic explanation for this distribution is that the experiencer argument is assumed to be in control of the implicit action expressed in the matrix clause. As such, the experiencer is more likely to exhibit subject-like properties, which include nominative case marking. However, this interpretation only extends to the variable clauses and not to conditional clauses, whose implicit propositional argument is assumed to express events that are beyond the experiencer's control. Morphosyntactic marking of the experiencer as subject is also observed by Allen for constructions with explicit propositional arguments in 27 out of 109 instances in Shakespearean usage. Both the NO PROP construction and the PROP construction thus differ in their permission of a subject option from the equivalent 2NP constructions with two nominal arguments, which do not develop such an option. It is not entirely clear from Allen's presentation, however, whether the subject option in these cases should be taken to refer primarily to the preverbal position of the experiencer, or whether it is also reflected in unambiguous instances of morphological case marking. The control of the experiencer, at any rate, is thought to be indicated by the fact that the propositional argument typically takes the form of an infinitival clause, in which the experiencer assumes the role of an understood (PRO) subject. While the overwhelming realisation of the propositional argument by a non-finite clause and the associated control of the experiencer clearly does not constitute a sufficient condition for subject assignment, the cases in which such assignment takes place are numerous enough, according to Allen, to indicate a tendency that can be explained in semantic terms.

The overall development of impersonal verbs and the demise of preverbal oblique experiencers is described by Allen as a gradual loss of lexical case marking in favour of structural case marking, coupled with an increasing tendency to assign the role of subject to the highest-ranking semantic role in the case of type-ii verbs (cf. Allen 1995: 284–286, 291–326). While she does not view the loss of morphological distinctions as an immediate trigger of this development, she does identify two morphological factors that weakened the system of lexical case assignment. First, the syncretism of dative and accusative case, which Allen dates to the period between the mid-12th and the early 14th century, eliminated the option of lexical case

marking for objects and resulted in the emergence of direct instead of indirect passives of monotransitive verbs that had previously governed a non-accusative object, although some of the direct passives need to be assumed to have been borrowed into certain dialects, since these retained the case distinction longer. More important for the present context, however, is the fact that this development restricted the scope of lexical case marking to subjects. Second, the loss of the ability of full NPs to bear lexical case marking, an analysis which Allen prefers over the assumption of covert case marking in order to account for a perceived lack of verb agreement with these NPs (cf. Allen 1995: 232–236), provided evidence to the language learner that lexical case marking did not always apply to experiencer arguments, which could be extended to pronouns by less conservative speakers. The presumed inevitability of the process due to systematic pressure towards uniform structural case marking is reminiscent of von Seeffranz-Montag's assumption of a cross-linguistic typological drift towards preverbal nominative subjects, although the present scenario invokes a more specific large-scale analogy within English, by which the more productive personal pattern is ultimately generalised. Reanalysis or, in different terms, a change in parameter settings, is thus viewed not as the initial trigger but as the final stage of the diachronic process, in which lexical case marking, which had become increasingly optional, is finally abandoned in favour of general non-application. The steps towards this abandonment involve a gradual disfavouring of lexical case marking, which proceeded for individual lexical items by the increasing failure to apply this kind of case marking to experiencer arguments in lexical frames which stated lexical case marking as an option, and across lexical items by the introduction of optionality into lexical frames in which it had previously been obligatory. The last constructions with oblique experiencers in Allen's investigation of 16th-century authors involve verbs like *thinken* 'to think, seem', *semen* 'to seem' and *listen* 'to desire' in their use with clausal complements as well as fossilised expressions such as *methinks*, which require very specific lexical entries (cf. Allen 1995: 279–284). While this may seem surprising at first, since instances of the NO PROP and the PROP construction of *plesen* were previously analysed as examples of a semantically motivated introduction of nominative experiencers, Allen identifies communicative factors like the acceptability of a proposition for the addressee and deference of the speaker in making polite requests as additional factors in the distribution of oblique and nominative experiencers (cf. Allen 1995: 342f.). The gradual disfavouring of lexical case marking, which ultimately led to the generalisation of structural case marking, can thus be viewed as a process that was at least partly constrained by semantic and communicative factors.

2.2.1.2.2. Criticism

While the incorporation of lexical entries with optional case assignment is not specific to Allen's account, her assumption of a process of gradual disfavouring that affects different lexical items at different times and to different degrees distinguishes it not only from the initial proposal of a cataclysmic and sweeping change in terms of a low-level reanalysis but also from Fischer/van der Leek's proposal of a general loss of the optionality of lexical case marking, which is assumed to have been previously available to all impersonal verbs in equal measure (cf. Fischer/van der Leek 1983, who later revise their proposal in terms of adverbial case, which is assigned to oblique constituents with non-argument status, cf. Fischer/van der Leek 1987). By its very nature, this account is much more suited to the incorporation of intra-lexical and inter-lexical variation, which appears to be a better representation of the empirical evidence than the prediction of reanalysis, which, at the very least, requires additional assumptions. The notion of gradual disfavouring is thus not subject to the same criticisms that were discussed above, but the relative lack of restrictiveness may, instead, be viewed as a drawback. There are, of course, ample means of providing a structural analysis of both personal and impersonal verbs in the framework of Lexical Functional Grammar, but the general principles and diachronically variable lexical specifications of the feature assignment to semantic roles and grammatical relations do not, in themselves, predict the observed diachronic changes, they merely describe them. There may be certain generalisations that make one version of the theoretical framework more attractive than another, but there also appears to be some flexibility in the application of such generalisations, which weakens their descriptive and explanatory force. The double-object construction with two oblique arguments, for example, is analysed by Allen as the regular assignment of lexical case to the subjects in combination with structural case assignment to objects, but this assumption requires the suspension of the case hierarchy as one of the general principles in accounting for regular case patterns, since this hierarchy would otherwise require the assignment of nominative case to the object, similar to the relevant instances of type-ii verbs (cf. Allen 1995: 240f.).

A second point is that the comprehensive discussion of textual examples, which certainly presents one of the merits of Allen's study, touches upon several points of evidence that do not support the theoretical assumptions and thus need to be accommodated by alternative explanations. One such example relates to Allen's account of syntactically possible case frames, which is intended to capture not only the attested patterns of personal and impersonal verbs but also the predictions about patterns that are impossible according to theoretical expectations. Specifically, the occurrence of patterns with a dative experiencer and an accusative stimulus is

dismissed by Allen in preference of alternative explanations in terms of early phonological attrition, but this preference seems to be at least partly related to the fact that accusative case marking of the stimulus argument would present counter evidence to the case hierarchy, which Allen assumes to be in place during the relevant period of Old English. Similarly, her assumption that the lexical case marking of subjects is not generally optional for impersonal verbs is justified by the absence of transitive constructions with *ofhreowan*, but this entails that the evidence of such constructions with *hreowan* either needs to be regarded as not genuine or as a lexically motivated exception. The existence of potential counter examples that require some sort of accommodation is, of course, not a very uncommon situation in historical linguistics, and the fact that such examples are discussed and not simply suppressed by Allen is certainly commendable. Depending on how closely one follows her argumentation, however, the impression remains that the formalism of Lexical Functional Grammar, although well-suited to the description of verbal argument structure, is not entirely comprehensive and that even a relatively detailed lexical account with only few generalisations is not able to account for the complete range of patterns that are attested during the OE period.

A different criticism relates to the learnability of increasing optionality and its mental representation, which, according to Lightfoot (1991: 126f.), would require an amount of lexical specification that is psychologically implausible and thus renders lexical accounts without any explanatory force. The full scope of this discussion need not be entered into here, since the aim is not to prove or disprove the reality of lexical frames with optional case assignment, although the question of the extent to which statistical information about the frequency of a given construction affects the mental representation of language during the initial period of acquisition as well as beyond that period is certainly an interesting one. The main point in the present context, however, is that a diachronic account of impersonal verbs should ideally strike a balance between a maximally comprehensive account in terms of individual changes in lexical specifications and a more principled parsimonious account, which appears to be only superficially more attractive in ignoring much of the attested variability of the verbs in question. A theoretical account in terms of changing lexical features alone does not seem to constitute an explanation of the change, while a transformational account in terms of structurally required reanalysis constitutes a claim that is not only heavily dependent on the specific syntactic theory but also stronger than the empirical evidence permits. A more promising explanatory device seems to be the mechanism of analogy that is inherent in Allen's assumption of systematic pressure towards structural case marking. Analogy is a well-established cognitive process and also a frequently attested mechanism in diachronic language change. The assumption of a

general analogy of impersonal verbs according the pattern of personal verbs is, of course, not a very specific explanation. The question is, therefore, whether more local analogies within or between verbs of a given semantic class can be observed. Likewise, Allen's semantic explanation of at least some of the instances of early introductions of nominative experiencers as well as the functional characterisation of communicative acts as deferential and directed at gaining the interlocutor's acceptance can be regarded as more informative explanatory devices, since they are extrinsic to the mere reformulation of rules that govern the well-formedness of a language's syntax. Semantic and functional influences on the development of impersonal verbs thus seem to constitute aspects that warrant further investigation.

2.2.1.3. Principles and Parameters

Also Lightfoot eventually abandoned both the Transparency Principle and the Trace Erasure Principle as explanatory devices for the reanalysis of oblique experiencers and attempted to provide an alternative account which incorporated, or at least permitted, the apparent gradualness of the diachronic development of impersonal verbs in the form of intra-individual diglossia (cf. Lightfoot 1991: 128–141). The primary motivation for this assumption is developed from the underlying framework of Principles and Parameters, which stipulates discrete parameter settings that complement a set of basic principles within a presumably innate Universal Grammar. On account of their discrete nature, the parameter settings, which are intended to account for the observable cross-linguistic differences in linguistic structures, are not particularly well suited to describing tendential changes in the lexical distribution of optionality or statistical frequency. Instead, parameter resettings are taken to result in the kind of cataclysmic change that is much more akin to the mechanism of reanalysis based on syntactic principles. An important question that arises in this context is, of course, the cause of the parameter resetting itself. While the different settings that occur during the initial process of language acquisition can be accounted for by the assumption that certain parameter settings present unmarked options that take effect by default while the corresponding marked options require positive evidence to be set, a diachronic change in these settings presupposes some kind of preceding change in the linguistic input that constitutes the triggering experience, since identical inputs should typically, even if not necessarily, result in identical output grammars. Part of Lightfoot's study is thus devoted to establishing the nature of the triggering experience, which he considers to be largely restricted to the evidence of matrix clauses with so-called "degree-0 complexity" or, more generally, to material from unembedded binding domains,

which include certain elements from embedded clauses (cf. Lightfoot 1991: 31f.). In addition, the evidence must be robust, i.e. frequent enough to have an effect on the triggering experience. The changes that eventually alter the triggering experience substantially enough to result in a parameter resetting are merely regarded as changes in the application of optional rules, which do not constitute grammar changes but rather a kind of accidental variation, whose cumulative effect may nevertheless lead to a parameter resetting (cf. Lightfoot 1991: 67f.). Changes in the application of verb movement and particle placement in matrix clauses, for example, are interpreted as causing a parameter resetting of basic word order from OV to VO, which is then extended abruptly to embedded clauses as a direct consequence of this resetting.

With regard to impersonal verbs, Lightfoot adopts the fundamental distinction between structural and lexical case assignment or, in different terms, between structural cases assigned at the level of S-structure and inherent cases assigned at underlying D-structure (on the distinction of these cases cf. Lightfoot 1991: 111–115). He rejects the idea, however, that optional assignment of inherent cases could be a lexically specified property and instead assumes that it constitutes a (marked) parameter setting that affects all pertinent verbs in equal measure. The apparent restriction of this optionality to impersonal verbs is motivated by their classification as a distinct subgroup of psych-verbs, which express psychological or physiological states. This definition, indeed, includes most of the one- and two-argument experiencer verbs typically considered in the context of impersonal constructions, although existential experiences denoted by verbs of happenstance as well as the situations denoted by verbs of obligation and necessity do not fit the definition equally well. Since the optionality of inherent case assignment is a parameter setting, its loss, which is now seen as the result of a gradual loss of the morphological dative category, must affect all of the relevant verbs simultaneously. The continued variation between oblique experiencers and nominative experiencers is therefore relegated to a diglossic scenario, in which two simple grammars exist side by side, one of which contains lexical entries that specify inherent cases and one of which does not. The assumption of different grammars potentially accounts for variation within a speech community and could thus explain the textual evidence of a continued occurrence of oblique experiencers besides nominative experiencers. Lightfoot also needs to assume, however, that grammars with and without inherent cases were present in individual authors or scribes, implying a kind of internal or intra-individual diglossia, whose limitations, as Lightfoot remarks, are not particularly well understood (cf. Lightfoot 1991: 138). As such, the diglossic scenario eliminates much of what the initial proposal of a loss of the distinction between structural and inherent cases is supposed to explain. The actual variation between different

construction types is neither represented in terms of the linguistic system nor accounted for in terms of a diachronic explanation. The acknowledgement of a gradual diachronic process thus presents an objective improvement over Lightfoot's previous accounts, but the continued rejection of lexical differentiation leaves much to be desired in terms of an explanation, which the theoretical construct of a diglossic scenario is not able to provide.

The notion of parametric change as pervasive mechanism in diachronic syntax is continued by Roberts, who discusses reanalysis as one of the principal types of syntactic change within his adopted version of Minimalism (cf. Roberts 2007: 122–141). Assuming that parameter settings can impact a range of clustered grammatical properties simultaneously, he regards reanalysis as symptom of a preceding parameter change, which can trigger several reanalyses at once. The mechanism of reanalysis is thus ultimately reducible to parametric change and, like earlier accounts of reanalysis, conceived of as an essentially instantaneous change that takes place during the transmission of individual grammars between members of different generations. An interesting question in this context is the one raised above about the nature and the order of changes leading up to or resulting from a parameter change. Differences in an acquired grammar G_2 compared to a parental grammar G_1 seem to presuppose a difference in the input of these two grammars, but this raises the question of why users of G_1 should produce an output that differs from the one on the basis of which their own grammar was formed. Two principal ways are proposed by Roberts to account for such differences, viz. external language contact or subtle internal changes in extrasyntactic domains. An example of the latter case is provided by the French question particle *ti*, which arises in colloquial language by reanalysis of the third-person singular masculine pronoun *il* 'he' in inversion with *a-t* 'has' and under deletion of word-final *l* (cf. Roberts 2007: 129–132). This account of a reanalysis of morphological boundaries is, of course, quite plausible, although it also shifts the need for an explanation to the domain of phonology, since the conditions of segmental deletion, which provides the basis for reanalysis, as well as the frequency required of this phenomenon in order to trigger reanalysis, remain open. If statistical shifts in the frequency of such phonological processes, or in the frequency of syntactic processes like verb movement and particle placement as discussed by Lightfoot, if such shifts are regarded as essentially accidental variation, they would, perhaps, neither require nor afford any further explanation. This is basically the opinion expressed by Lightfoot in his version of parametric change. An entirely different option is to regard changes in surface structures as the result rather than the cause of reanalysis, in which case the changes themselves would be accounted for, but not the reanalysis, if it cannot be attributed to other grammatical changes, which would then require an explanation of their own.

This scenario corresponds to Lightfoot's earlier accounts of impersonal verbs, in which the deep-structure reanalysis of oblique preverbal experiencers is manifested more or less gradually following the establishment of base-generated SVO.

In Roberts' own account, the changes to impersonal verbs are no longer viewed as the result of a single low-level reanalysis with large-scale repercussions nor as the result of a parametric abolishment of lexical case after the gradual loss of morphological dative but, following Allen, treated as a case of lexical diffusion. Robert bases his view on Allen's findings that impersonal verbs exhibit individual, although not necessarily entirely piecemeal, changes in lexical properties over a long period of time, which he proposes to conceptualise as a series of small-scale lexical reanalyses. In addition, he considers the option that the process of lexical diffusion may be contingent upon extra-linguistic social or historical factors (cf. Roberts 2007: 152–161, 297–300). Both of these attributes are seemingly contrary to the assumed catastrophic and language-internal nature of parametric change, but if lexical diffusion, which is taken to affect not only lexical properties like the argument structure of impersonal verbs but also the properties of functional items, if this process is essentially regarded as a series of discrete reanalyses, the superficial impression of gradualness resulting from a series of such changes can be reconciled with the principal assumption of syntactic change as instantaneous. With regard to each individual lexical reanalysis, the purely formal account is, of course, much less powerful than the more far-reaching application of parametric change to functional items. While it provides a way of integrating the actuation of changes in either lexical or functional categories with Roberts' overall account of parameter change, the implementation of such changes within a speech community and the potential involvement of demographic factors and aspects of register in this process appear to be of equal if not greater importance for a proper understanding of syntactic change. This applies not only to its social diffusion but also to the variation between different argument structures of a single lexical item that is observable in the use of individual speakers. In contrast to Lightfoot, who delegates the variation between oblique and nominative experiencers to supposedly ill-understood limitations on intra-individual diglossia, Roberts assumes formal optionality to be a grammatically available feature, at least for syntactic operations like the scope of movement (cf. Roberts 2007: 307–309). The presence of "true optionality", according to Roberts, implies that no semantic, demographic or functional features account for the distribution of a given set of variants. If one of these variants takes on the value of a social identity, for example, this may lead to its propagation and ultimately to the replacement of other variants in the form of syntactic change, provided that structural reasons and markedness considerations do not oppose (cf. Roberts 2007: 332). Although not explicit in

Roberts' account, it seems possible to extend this explanation to the diachronic reorganisation and the attendant synchronic variation of case marking specified in the lexical entries of impersonal verbs, or, formulated in terms of a construction-based approach, to the changing application of constructional templates to impersonal verbs. The idea that extrasyntactic factors are involved in the spread of nominative experiencers, for both individual lexical items and across different items, will be elaborated in more detail in the following section.

2.2.2. Semantic Hypothesis

The discussion of earlier accounts of the diachronic development of impersonal verbs has illustrated that both the actuation of the changes that occurred and their implementation with individual verbs and across lexical items can be conceptualised quite differently, depending on the adopted framework of synchronic syntax. The assumed explanatory power of a single catastrophic system change in the form of a structurally required low-level reanalysis or a parameter resetting that affects the availability of lexical case marking is offset by its apparent inadequacy to account for the empirical evidence of a gradual and lexically diverse process that includes periods of renewed productivity. Additional explanations are required for both the period preceding such a change and for the period that follows, since the assumption of a catastrophic system change, in which the earlier analysis is either prohibited under the strong view of Lightfoot's earlier accounts or at least not reinstated due to parameter ambiguity in the more moderate account of Principles and Parameters, begs the question of how the conditions for such a change in the triggering experience arose, not least because they are assumed to have developed within the limitations of an inter-generational language transfer. The period following the assumed reanalysis is of interest, since it is now generally acknowledged that no sweeping change of impersonal verbs occurred at any given point in time. The assumption of speaker-internal diglossia in order to account for the continued variation in the case marking of experiencer arguments after the assumed abolishment of lexical case marking appears to be a makeshift attempt at reconciling the observed variation with a conceptualisation of language as a minimally complex system of discrete structures, which lacks explanatory power in important areas, however, if the conditions on such diglossia are left entirely open. Roberts' assumption of optionality as a grammatically viable option integrates variation into the formal account of syntax in some areas, at least, but the analysis of the lexical diffusion of nominative case marking of experiencer arguments into a series of discrete reanalyses of individual lexical entries, while consistent with his general framework, is certainly less revealing in terms of a

purely formal description and leaves room for a more detailed account in terms of extrasyntactic factors. Many of the attendant problems of synchronic variation and gradual change also appear to be quite specific to the formalisation of syntax in the generativist tradition, which places great emphasis on grammatical well-formedness, a concept that is typically regarded as a binary opposition. A usage-based view of language with its recognition of the changeable and fuzzy nature of linguistic entities does not meet with the same problems, and the relation of these entities to semantic and functional aspects shows a much greater appreciation of language as a tool for communication. The basic assumption that semantic factors may influence the development of impersonal verbs is already present in Allen's account. Overall, her study has a predominantly formal orientation. Its principal aim lies in the faithful investigation of the gradual and lexically diverse pathways which the affected impersonal verbs take during their development according to the empirical evidence as well as in the formal representation of their lexical entries in terms of Lexical Functional Grammar, which incorporates the assumption that subject status could be assigned to oblique constituents. The study also involves the semantically motivated assumption, however, that nominative case was extended early in situations in which the experiencer could be conceived of as a responsible and thus more agent-like participant. The more general notion that the extension of an innovative form or construction constitutes a grammatically ordered process was discussed in some detail by Timberlake, who regarded this process as a direct complement to reanalysis. While his assumption that changes in syntax take the form of an instantaneous reanalysis does not constitute a prerequisite to an investigation of the extension of such changes across different contexts, it is still worth considering how this process was first envisaged by Timberlake in his seminal contribution to the discussion of reanalysis and actualisation, whose basic reasoning will be applied to the development of impersonal verbs in the following section.

2.2.2.1. Underlying Theoretical Concepts

2.2.2.1.1. Actualisation

The actualisation of reanalysis has been described by Timberlake (1977) as the extension of grammatical changes across different surface contexts. His account relies on the common distinction of transformational grammar between deep structure and surface structure, in addition to which a distinction between grammatical rules proper and a set of usage rules is made, which mediate between the two levels of grammar. As in the traditional account, the starting point of reanalysis and its actualisation lies in structural ambiguity, which is thought to

arise in limited surface contexts from phonological or morphological changes. Structural ambiguity provides the basis for a grammatical reanalysis of the deep structure, and this innovative analysis is then gradually extended into morphosyntactically unambiguous contexts. The apparent time gap between reanalysis and its ultimate manifestation across all applicable surface structures is accounted for by the assumption of additional subrules, which language users maintain in order to reconcile their innovative grammar with the conflicting output of the parental generation. This leads to the provisional preservation of constructions predicted by the earlier grammar, even though the innovative grammar is already acquired by such speakers. Crucially, the process of actualisation is considered to be grammatically constrained and thus susceptible to a theoretical account, which may be informed by theoretical generalisations as well as empirical observations.

Two examples of actualisation are discussed in more detail by Timberlake, both of which involve the assumption of implicational hierarchies. They concern the replacement of variable case assignment to constituents in subject-to-object raising constructions by regular assignment of genitive case in Finnish, and the replacement of genitive case by accusative case for objects in negative clauses and lexically governed objects in Russian. The former development is described as the result of a phonological merger of word-final **-m* and **-n*, which led to a syncretism of accusative and genitive case in singular nouns. After the merger, such nouns could be analysed as objects of the matrix verb, which receive accusative case in active clauses, or as subjects of a participial clause, which receive genitive case. According to Timberlake, this situation caused a low-level reanalysis, whose actualisation appears to have affected pronouns earlier than plural nouns. An early innovative example in the 16th-century writings of Agricola involves a pronoun, while conservative grammatical treatments of Finnish exemplify the older construction with the help of (inanimate) plural nouns. This is, of course, a very slight empirical basis, but it illustrates the basic point that the directionality of actualisation can be constrained by linguistic hierarchies such as the one in (57):

(57) pronouns > animate nouns > inanimate nouns

The point is reinforced by Timberlake (1977: 154–156) by drawing on a related change that took place in Finnish infinitival clauses. Similar to the situation in participial clauses, former raising verbs such as *antaa* ‘to let’, *sallia* ‘to permit’ and *käskää* ‘to order’ formerly assigned variable case to their objects, depending on the type of construction. For pronouns, this was replaced with genitive case, which is the typical case of subjects in non-finite clauses. For nouns, however, an optional choice between accusative and genitive case is reported in active

clauses if the noun is non-agentive. Native-speaker judgements suggest that the option of accusative case marking was already falling out of favour at the time of Timberlake's inquiries, but for the period leading up to the abolishment of accusative case, a hierarchy that partly resembles that of participial clauses can nevertheless be established, cf. (58):

(58) pronouns > agentive nouns > non-agentive nouns

An important point is that these hierarchies, despite the overlap in the preceding examples, are taken to be specific to the type of change to which they apply. In the cases above, the new construction specifically marks the constituents as subjects of non-finite clauses. Actualisation is therefore expected to affect constituents with subject-like properties earlier than other constituents, since the innovative genitive case is the unmarked case of infinitival subjects. Semantic properties like animacy and agentivity, the latter of which appears to be defined largely in terms of volition judging by the examples that Timberlake provides, appear to be amenable to such an interpretation.

A differently motivated set of hierarchies is proposed by Timberlake for the change in case marking of objects in negative clauses and lexically governed objects in Russian. The basic function of the case under replacement is defined as a kind of quantitative limitation of the object's participation in the verbal action, which is, perhaps, most apparent in the partitive use of the genitive, but which can also be perceived in its other uses. In negative clauses, the limitation is complete in the sense that the object is entirely excluded from the event, and in the case of lexical government, the object itself presents a kind of circumstantial limitation on the event, while it does not immediately participate in it. In as far as quantification plays a role, the use of the innovative accusative, which does not express quantification, would be expected to occur more readily in contexts of highly individuated entities, while the conservative genitive case would be expected to be retained longer with quantifiable entities. This seems to be reflected in the reported acceptability ratings of these cases depending on different contexts in negative clauses, on the basis of which the hierarchies in (59) can be established:

(59) proper noun > common noun

human > animate > inanimate

concrete > abstract

singular > plural

definite > indefinite

In each case, the higher-ranking item provides the context in which the innovative accusative is either obligatory or more acceptable than the genitive. Similar hierarchies can be illustrated for the replacement of lexically governed genitive by accusative case. It is not quite clear from Timberlake's examples, however, what the basis of the semantic interpretation of the examples and their associated acceptability ratings is. The latter might be derived from grammar books, linguistic surveys, individual informants or from the author's intuition. The former may be fairly straightforward for most of the above distinctions, but since definiteness is not marked by the use of articles in Russian, the interpretation of NPs such as those in (60), Timberlake's (27), where it is not unambiguously inferable from verbal aspect, seems to depend on an unspecified context or, in light of the proposed hierarchy of definiteness, on the nature of case assignment, in which case the interpretation would be circular:

(60) Ja ne našel cvety / ?cvetov.

I not found the-flowers(ACC) / flowers(GEN)

'I did not find the / any flowers.'

The problem is less pronounced for the other hierarchies, since proper nouns will generally be distinguishable from common nouns, animate nouns from inanimate nouns, and singular nouns from plural nouns, although the difference between concrete and abstract nouns may be less clear-cut. Similarly, pronominal items are usually distinguishable from nominal items, but the applicability of an agentive or volitional interpretation, which is supposed to affect the spread of genitive subjects in Finnish infinitival clauses, is a matter of judgement, which, ideally, should be based on criteria that are independent of the proposed hypothesis. An example of this situation is provided in (61), which is a conflation of Timberlake's examples (16) and (17):

(61) Annoin niiden miesten / ne miehet jäädä reserviksi.

let those men(GEN) / those men(ACC) stay reserves

'I let those men stay in the reserves (as they wished / without regard for their desires)'

While different degrees of volition may, indeed, be applicable to the object of the matrix clause in such cases, the identification of the applicable interpretation seems to depend on non-objectifiable criteria, a fact which is less problematic in the case of living languages, for which native-speaker judgements can be obtained, but which presents a significant problem for investigations of historical data, which, apart from relatively rare instances of grammatical treatises, rely on the interpretation of the researcher. Nevertheless, the basic claim that the

extension of innovative linguistic features, which, in Timberlake's context, is regarded as the actualisation of a reanalysis of deep structure, may proceed along specifiable pathways like those suggested above is worthy of further consideration.

Before proceeding with a characterisation of potential conditions on the development of impersonal verbs, it should be noted that Timberlake's notion of syntactic change as a grammatically ordered process does not necessarily presuppose the occurrence of reanalysis in order to be of interest, even though the term actualisation is most appropriate when this is the case. In their comprehensive treatment of historical syntax, Harris/Campbell (1995) describe reanalysis and extension as two principal mechanisms of syntactic change besides language contact, but while actualisation and extension are conceptually similar and partly overlapping processes, extension is considered to be generally independent of reanalysis. The common definition of the latter mechanism as a process that affects underlying linguistic structures without immediate surface modifications is supported by the authors, but there are at least two important differences in their account compared to the ones discussed previously. First, structural ambiguity, but not necessarily opacity, is seen as a prerequisite for reanalysis (cf. Harris/Campbell 1995: 70–72). That is, while a given surface structure must be open to at least two different interpretations, it is not a requirement that both of these are pre-existent in the grammar or that the traditional interpretation becomes unclear and thus unavailable after reanalysis. Reanalysis is thus thought to include the emergence of innovative interpretations, and these may coexist with earlier interpretations rather than immediately replace them (on the term “neoanalysis” rather than reanalysis for the emergence of innovative interpretations cf. also Traugott 2015: 63 with reference to Andersen 2001[c]: 231³). The latter point is important, since it implies that a process of gradual extension with multiple coexisting interpretations will be the rule for changes by reanalysis rather than an exceptional case. The issue of its synchronic representation in terms of speaker-internal diglossia or formal optionality, which is a point of concern for Lightfoot and Roberts, is, of course, not resolved by this assumption. Regarding the second mechanism, Harris/Campbell maintain that actualisation can often be regarded as an instance of extension, although the two concepts are not coextensive in that actualisation may involve additional reanalyses besides extension, and extension itself need not be the result of reanalysis, but may also occur as part of a contact-induced change (cf. Harris/Campbell 1995: 80f.). The general mechanism of extension is essentially defined as the converse of reanalysis, viz. as a change in the surface manifestation of a syntactic structure that does not affect the underlying structure. Crucially, this mechanism is considered to be systematic and specifically constrained by the requirement that old and new contexts of a syntactic rule under extension

must combine into a natural class with respect to categories that are already relevant to this rule (cf. Harris/Campbell 1995: 101f.). De Smet (2012: 606f.) correctly remarks that this formulation of the constraint requires the additional definition of a natural class with respect to a given rule change, a point which Harris/Campbell appear to establish largely by inference from their empirical data. With regard to impersonal verbs, or “inversion verbs” in their terminology, the process of extension is conceived of as a lexical extension, which leads to changes in the lexical features of the affected verbs. Such changes involving lexically determined rules are considered potential exceptions to the general formulation of the constraint on extension in that not all members of a natural class are necessarily included during each step of the extension process (cf. Harris/Campbell 1995: 108–115). This renders the question of what constitutes a natural class in the context of the extension of nominative case to experiencer arguments a moot point, although it seems reasonable to assume that semantic subclasses like verbs of emotion and verbs of cognition may play a role. A different line of argument involves Timberlake’s concept of change according to markedness principles. This reasoning will be explored further in the following subsection.

2.2.2.1.2. Markedness

The notion of markedness as the ordering principle behind the grammatically ordered manifestation of diachronic change in language use has been explored in more detail by a variety of scholars. A notable contribution is made by Andersen (cf. Andersen 2001a, b, c), who regards the process of actualisation according to markedness as an essential part of his speaker-oriented theory of language change (cf. Andersen 2001c: 225). On a fundamental level, he argues that markedness is a domain-general cognitive principle, which allows for the deductive explanation of both synchronic and diachronic empirical observations rather than requiring inference from empirical properties such as frequency or different degrees of allophonic or allomorphic variation, which he considers to have been mistakenly applied by many of his colleagues. The examples of oppositional pairs of unmarked and marked terms which he adduces from a range of semiotic systems including ritual, fiction and various levels of linguistic expression do not appear to be immediately deducible from any single theoretical concept, however, especially since the more general instantiations of markedness are considered to be potentially specific to a given culture or period (cf. Andersen 2001b: 22–37). This means that the classification of a set of given terms as either unmarked or marked essentially continues to be based on the patterns that emerge from empirical observations, quite

contrary to the author's statement. One of his examples involves Timberlake's hierarchical organisation of grammatical contexts relevant to the replacement of genitive objects in Russian, which is given in (59) above. The alignment of the opposing terms with different degrees of perceived individuation appears to be a feature specific to the change in question, however, and not deducible from a general concept of markedness as such. A change-specific motivation of how a given change is expected to unfold according to markedness principles thus remains necessary in order to avoid circularity of the argumentation.

Within Andersen's general theory of change, an additional distinction is made between internally motivated "evolutive" changes and externally or "pragmatically" motivated changes (cf. Andersen 2001b: 33f.). This distinction is relevant because it affects the direction that actualisation is expected to take according to markedness. While Timberlake's earlier discussion is limited to the former type of changes, which involve an earlier or more consistent spread of innovative features in unmarked contexts, the latter type of changes is taken to involve a marked innovative variant, which manifests itself earlier in marked contexts. Specifically, Andersen distinguishes between different change scenarios, each of which include a series of partly identical subchanges. Coinage, remedial change and borrowing are categorised as pragmatically motivated scenarios, in which an individual speaker either draws on their own grammatical competence for expressive or therapeutic purposes or on another linguistic tradition with the explicit intention of fulfilling a particular communicative need. Extension, transference and interference, on the other hand, are considered to be less intentional transfers of existing linguistic structures into new contexts, either language-internally or in situations of bilingualism or language contact. In each of these change scenarios, the innovative feature may be subsequently adopted by other speakers, actualised through manifestation in language usage and ultimately reanalysed as an integral part of grammar during the process of language acquisition by a new cohort of speakers (cf. Andersen 2001c: 229–231). Actualisation, in this framework, is thus essentially equivalent to the manifestation of grammatical features in language usage, and its distinction from language usage in general depends mainly on the fact that it involves innovative grammatical features which are part of the process of diachronic change. Reanalysis, on the other hand, refers to the general process of integrating such innovative features into grammar during language acquisition. These do not need to be restricted to innovations arising from structural ambiguity, although this particular kind of reanalysis, which thus receives a double function as a factor in both the actuation and the implementation of change, is considered to be a major source of grammatical innovations that bring about evolutive language change (cf. Andersen 2001c: 233).

With regard to impersonal verbs, the most natural assumption seems to be that the extension of nominative experiencers constitutes an evolutive change, which is actualised earlier and more consistently in contexts in which the innovative nominative case marking is unmarked. This need not be true, however, if borrowing or other kinds of contact phenomena are involved. A speaker's competence in Anglo-Norman French, for example, may have provided models of transitive or reflexive constructions with nominative experiencer arguments, which could have been imitated in a pragmatically motivated change scenario. The productive extension of oblique experiencers to both borrowed and indigenous verbs shows that the changes to impersonal verbs are generally quite complex. Even if evolutive change is taken to be the most central scenario, the theory of markedness, in the form in which it was outlined above, does not offer much help in explaining the way in which the change unfolds. While markedness may well be a valid cognitive principle, the definition of unmarked and marked contexts in relation to a specific change continues to require an independent motivation, which entails the risk of circularity, since observed patterns may provide the basis for a classification according to markedness, which can then be used to explain the observed patterns (cf. Smith 2001: 205 and De Smet 2012: 606 for similar remarks). Under the present hypothesis that the extension of nominative case to experiencer arguments of impersonal verbs may have been susceptible to their semantic interpretation as more or less agentive, grammatically unmarked contexts would have to be motivated by the degree to which they correspond to these interpretations. The notion of linguistic change according to markedness principles offers no specific restrictions of this kind, however, apart from the general assumption that both evolutive and pragmatically motivated change take the form of a grammatically ordered process. The main attraction of Timberlake's and Andersen's discussions of markedness pairs for the present context thus lies in the fact that the suggested grammatical distinctions between singular and plural forms or between pronouns and full nouns offer relatively objective criteria that are immediately observable in the historical records of a given language and do not require native-speaker information or personal judgements by the researcher for disambiguation. The identification of relevant criteria that reflect different degrees of agentivity or volition would certainly be much preferable over a potentially subjective interpretation of individual instances of experiencer arguments based on their respective discourse environment. The remaining sections of the theoretical part of this study will attempt to identify such criteria based on a wider discussion of semantic roles (section 2.2.2.1.3), which provides the theoretical background to the present conceptualisation of agentivity (section 2.2.2.1.4).

2.2.2.1.3. Semantic Roles

The relevance of semantic roles to the synchronic system of impersonal verbs is already implicit in the observation that the majority verbs that are commonly discussed under this heading can be subsumed under the heading of experiencer verbs. It is clear, however, that the term “experiencer” essentially constitutes a cover term in this context, since it is used to refer not only to arguments of verbs that express psychological experiences like emotion (e.g. *lician*), cognition (e.g. *bynčan*) and perception (e.g. *hyngrian*), but also to the relevant arguments of verbs of happenstance (e.g. *gelimpan*), obligation (e.g. *behofian*) and possession or lack (e.g. *wanten*), which optionally profile a participant that is affected by chance events or to whom a given moral or possessive relation pertains. It is conceivable that this range of more fine-grained semantic distinctions of the experiencer argument is reflected in the diachronic development of impersonal verbs as well. Similarly, distinctions of the second argument as expressing either the role of cause or the role of theme have already been used as a motivation for synchronic differences in case assignment to experiencer arguments as well as their diachronic acquisition of behavioural subject properties by Anderson (1986) and Allen (1995), which makes a more thorough investigation of the semantic features of the arguments of impersonal verbs a desirable endeavour. A fundamental issue that arises, however, is whether semantic roles need to be seen as lexically specified and thus constant across all instances of a lexical item, or whether they can be assumed to vary contextually and thus potentially differ between instances of a single verb. The latter seems to be implied by Anderson’s attempt to account for the variation between nominative and oblique experiencer arguments through attribution of different case relations (ergative and ablative) to the second argument of impersonal verbs in accordance with the observed case marking. Also Allen, while making a lexical distinction between two subclasses of type-ii verbs with a theme argument and Dative Object verbs with a cause argument, nevertheless acknowledges that this semantic classification does not apply equally to every occurrence of the representative verbs *lician* and *cweman*. There are thus two levels on which the influence of semantic features on the development of impersonal verbs can be investigated, viz. on the level of lexically distinct items with their appendant semantic frames and on the level of contextually distinct interpretations of these frames. In the former case, expectations about the diachronic development will be defined based on the semantic subclasses of impersonal verbs, while, in the latter case, additional criteria that motivate the contextual interpretation of a given semantic role will be required. In either case, the expected impact on the diachronic development of impersonal verbs would take the form of a statistical tendency rather than of absolute discrimination, since the roles of a given lexeme’s semantic frame would

need to be allowed to vary contextually, and since also the properties that influence the contextual interpretation probably do not facilitate a complete prediction of the distribution of nominative and oblique experiencer arguments.

One aspect that is frequently remarked upon in the discussion of theoretical frameworks involving the concept of semantic roles is the inherent difficulty of establishing a finite inventory of semantic roles, which appears to result from an apparent lack of precise criteria for their theoretical definition and empirical identification (cf. e.g. Jacob 2004: 105–109, Wotjak 2004: 10f.). According to the earliest proponents of such frameworks, the principal motivation for assuming semantic roles as an additional or primary level of linguistic description lies in their capacity to account for syntactic phenomena like control (cf. Jackendoff 1987: 369, with reference to Gruber 1965) or constraints on coordination (cf. Fillmore 1968: 22), which do not immediately follow from syntactic structure alone. Such contrastive contexts may, indeed, serve to illustrate the relevance of semantic roles for linguistic theory, but they are not functional as diagnostic tools for the distinction of semantic roles in a corpus study of historical data. Specific criteria in terms of an absolute definition of semantic roles are difficult to establish, however, if not even the scope and the basic types of semantic roles are theoretically agreed upon. The discrepancy between more or less extensive inventories of semantic roles like that compiled by Blake (1930), whose list of 87 locative and 26 other “case relationships” is probably cited as something of a curiosity by Dowty (1991: 548³), or that of six presumably innate “deep cases” initially offered by Fillmore (1968: 24f.) compared to the reduction of semantic roles to essentially two proto-roles of agent and patient (Dowty 1991) or two macroroles of actor and undergoer (Van Valin 2004) not only illustrates the varying levels of abstraction aimed for by the respective frameworks, but it also illustrates the fundamental opposition between conceptually motivated distinctions on the one hand and morphosyntactically relevant distinctions on the other. Since conceptually justifiable distinctions need not have morphosyntactic expressions in a given language, their identification as predictors in a diachronic change scenario would rest on an unstable and potentially contentious basis. A more promising approach thus lies in the more abstract notion of generalised proto-roles or macroroles. This concept implies that the identification of a given verbal argument’s semantic role can be based on its more or less prototypical fulfilment of the relevant criteria rather than on absolute agreement, and it entails a discussion of agent-like and patient-like properties, whose operationalisation in terms of grammatically manifested criteria may serve as a more robust diagnostic tool for an investigation of the diachronic changes under discussion.

A comparison of the two aforementioned theoretical frameworks that involve generalised semantic roles is provided by Kailuweit (2004), who contrasts the notion of feature-based proto-roles developed by Dowty (1991) and his successors with the concept of macroroles arising from logical structure developed as part of Role and Reference Grammar by Foley/Van Valin (1984) and continued by Van Valin/LaPolla (1997) and Van Valin (2004). The starting point of Dowty's proposal is the idea that the notorious difficulty of defining semantic roles in the traditional terms of discrete categories stems from the fact that they much rather constitute cluster concepts, in which the arguments of a predicate hold different degrees of membership (cf. Dowty 1991: 571f.). This membership is based on a set of predicate entailments, which arise from the lexically defined properties of a given argument and are distinct from utterance-specific entailments that arise from the properties of a given referent. For the role of proto-agent, the following contributing properties, repeated in a slightly adapted form in (62), are listed by Dowty (1991: 572):

- (62) a. volitional involvement in the event or state
- b. sentience and/or perception
- c. causation of an event or change of state in another participant
- d. movement relative to another participant
- e. existence independent of the event named by the predicate.

The last of these properties is given in brackets by Dowty, who considers independent existence a potential result of perspective-dependent discourse structure rather than part of the definition of proto-roles. The grammatical relevance of the properties of proto-agents, each of which contrasts with a correlating property of proto-patients, lies in their contribution to the selection of arguments for the roles of subject and object. The argument with the greatest number of proto-agent properties is expected to be lexicalised as subject, while the argument with the greatest number of proto-patient properties is expected to be lexicalised as direct object (cf. Dowty 1991: 576). Kailuweit remarks, in the context of argument selection, that the assumed process is not completely well-defined, since it is not clear whether all properties carry equal weight and whether the identification of arguments with either of the two proto-roles should essentially be conceived of as a matter of arithmetic calculation (cf. Kailuweit 2004: 85). This is particularly relevant in cases where two arguments possess an equal number of proto-agent and proto-patient properties. It seems, though, that Dowty had no absolute prediction of argument selection in mind, since the Corollary 1 to his Argument Selection Principle states

that, in cases of approximately equal entailments, either or both arguments may be lexicalised as subject (or object). Psychological predicates or “mental verbs”, which include verbs of emotion and cognition like the ones under discussion, are, in fact, discussed as a specific case of indeterminacy, since the experiencer of such verbs entails the property of sentience/perception, while the stimulus entails the property of causation (cf. Dowty 1991: 579f.). Both arguments are thus equally good candidates for identification with the role of proto-agent in terms of their agent properties. Interestingly, Dowty interprets the subject selection of so-called stimulus-subject verbs like PDE *please*, which lexicalise the stimulus argument as subject, to be congruent with an optionally available inchoative interpretation compared to the exclusively stative interpretation of experiencer-subject verbs like PDE *like* (cf. Dowty 1991: 580, 586f. with reference to a paper presented by Croft in 1986). According to Dowty, the change of state in the experiencer that is implied by this interpretation lends this argument an additional patient property, which motivates its selection as object. Compared to Allen’s motivation of subject formation by the distinction of volitional cause arguments and more neutral theme arguments, his emphasis is thus on the affectedness of the experiencer, although the variable interpretation of stimulus-subject verbs as stative or inchoative similarly implies the assumption of intra-lexical variation, despite the fact that Dowty’s properties are explicitly restricted to lexical entailments.

In Foley/Van Valin’s framework of functional syntax, the macrorole of actor is provisionally defined as “the argument which expresses the participant which performs, effects, instigates or controls the situation denoted by the predicate” (cf. Foley/Van Valin 1984: 29, cited by Kailuweit 2004: 96). While this characterisation is partly reminiscent of Dowty’s agent properties, it is not intended as a feature-based definition of actor, but as the highest level of semantic generalisation in a binary opposition between actor and undergoer as the two cardinal arguments of transitive predicates (cf. Van Valin 2004: 63). Below this level of semantic macroroles, two lower levels of generalisation are formally recognised, viz. verb-specific “semantic roles”, which include roles like giver, runner and killer, and “thematic relations”, which include the traditionally discussed concepts of agent, experiencer and recipient among others. The relationship between these three levels of semantic generalisation, to which a level of grammatical relations can be added, is suitably illustrated in the continuum of semantic roles provided by Van Valin (2004: 64), which is reproduced below in an adapted and slightly abbreviated form in figure 1:

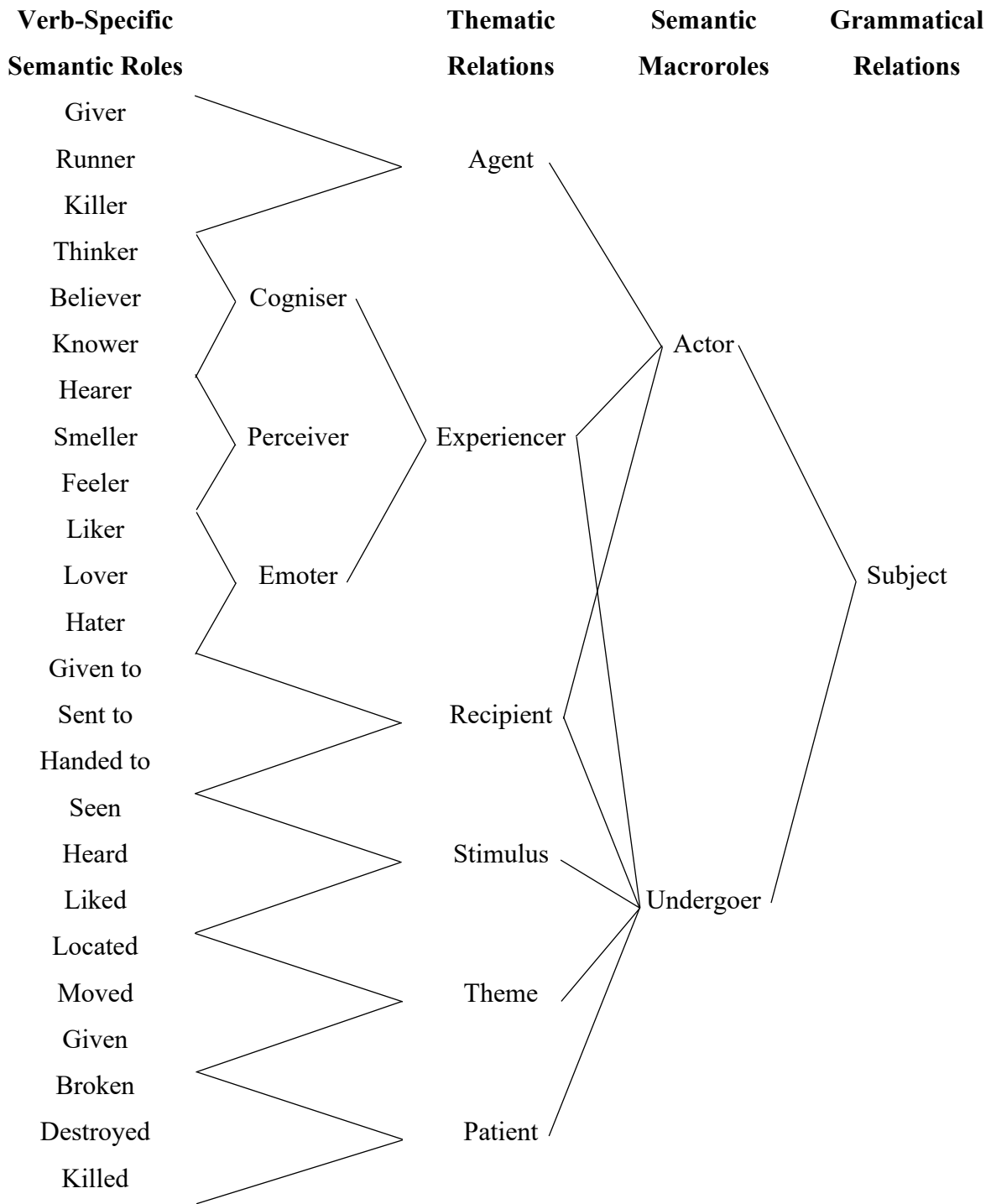


Figure 1 Continuum of semantic roles and grammatical relations (Van Valin 2004: 64).

The semantic and thematic relations in this continuum are arranged along two dimensions, the horizontal one of which specifies the degree of generalisation, while the vertical axis represents a semantic continuum of argument types, in which agents and patients constitute the prototypical endpoints (cf. Van Valin 2004: 69–71). The arrangement of the semantic roles and thematic relations on this axis reflects the degree to which they are taken to correspond to the prototypical endpoints of agent and patient in terms of activity and affectedness. The first argument of states involving perception, cognition and emotion, for example, is characterised

as less agent-like than the first argument of activity predicates, but on account of the presumed mental, emotional or perceptual internal activity of the relevant participant, it is characterised as more agent-like than the second argument of states, which is variously identified as the stimulus of perception, the content of cognition or the target of emotion (cf. Van Valin 2004: 70f., 68). As can be seen, the semantic argument types are additionally diversified in the area experiencer arguments, which are subdivided into cognisers, perceivers and emoters on a theoretically unspecified level in Van Valin's continuum. The arrangement of these subdivisions equally implies a higher degree of agentivity for the first argument of verbs of cognition than for that of verbs of perception and emotion. A more detailed analysis of the semantic role of so-called experiencer arguments may thus provide further insights for the development of impersonal verbs.

Two central conclusions follow from the preceding discussion of semantic roles. First, the experiencer argument of impersonal verbs can be analysed more precisely into a continuum of roles that more accurately reflect the semantic verbal classes, including verbs of cognition, perception and emotion, and these arguably exhibit different degrees of affiliation with the prototypical role of agent. The equivalent argument of verbs of happenstance, obligation and necessity, which do not presuppose any internal activity in the relevant participant, is further removed from the role of agent than all of the above. Second, the experiencer argument, in all of these cases, is located in the middle of the continuum between agent and patient and can therefore be conceptually and morphosyntactically subsumed under either of the prototypical endpoints in the spectrum. This circumstance is reflected theoretically in the relative indeterminacy of argument selection in Dowty's discussion of psychological verbs as well as in the variable linking of experiencers with the macroroles of actor and undergoer in the framework presented by Van Valin. Morphosyntactically, it seems to be realised in the non-nominative case marking of type-i constructions with a dative or accusative experiencer argument and a genitive stimulus argument. The selection of either of these arguments for nominative case marking arguably requires a decision of the language user in terms of identification with the prototypical role of agent. The question is, then, what kind of properties would entail a more agentive interpretation of either of the two arguments.

2.2.2.1.4. Agentivity

Dowty's list of proto-agent features, which was given in (62) above, provides a suitable starting point for the discussion of agentive properties, since it not only analyses the concept of

agentivity into several independent features but also presupposes that not all features need to be equally realised in the case of less prototypical agents. There are several limitations to the applicability of these features to the arguments of impersonal verbs, however. First of all, the feature of independent existence, which is regarded as a potentially perspective-dependent property by Dowty, seems to rule out effected participants from the concept of agent. A table made by a carpenter, for example, would presumably fall under this definition, but such a feature does not seem to be distinctive in the case of experiencer or stimulus arguments, both of which exist independently of the experience. This feature will therefore not be discussed further. Second, the feature of movement seems to be equally excluded from the situations that obtain in the case of impersonal verbs, at least if external movement is referred to. The following discussion will thus be limited to the three remaining features volition, sentience/perception and causation and their applicability to the arguments of the major semantic classes of impersonal verbs.

2.2.2.1.4.1. Volition

The first concept in Dowty's list of proto-agent features is volition, which has also been used as a criterion by Allen in her discussion of different types of experiencer arguments. There are at least two ways in which the concept of volition can be defined, one referring to a participant's favourable disposition towards the occurrence of a given event and one referring to the event's dependence on the participant's wilful execution (cf. also Primus 1999: 36). The latter definition will be taken as a basis in the following discussion. A typical instance of volition in this sense would be the intentional performance of an action as opposed to chance or accidental involvement. Intuitively, this definition does not seem to apply to the experiencer argument of verbs of emotion, however, since emotions are usually regarded as spontaneous events which cannot be controlled by the experiencer. Nevertheless, various distinctions of a more or less volitional construal of emotion have been discussed in the literature, and these can be subjected to several generalisations. For example, Pishwa notes that many oblique-experiencer verbs in Old English refer to negative experiences like fear, grief, suffering and doubt, which she considers to be more salient than positive experiences and thus to "give the sentient [entity] the feeling of total impact" (cf. Pishwa 1999: 133). Presumably, this is supposed to imply an overwhelming degree of affectedness on the part of the experiencer, who is thus less likely to control the emotional experience in the case of negative emotions. Positive emotions, on the other hand, would not affect the experiencer so much as to render them unable to consciously

process and thereby assume control of the emotion. A higher potential for an agent-like interpretation would thus be expected for the experiencer of positive emotions than for the experiencer of negative emotions, if this is, indeed, a valid distinction. An additional but related point lies in the fact that the experiencer may actively try to sustain positive emotions, while negative emotions will most likely not result in such efforts. This interpretation seems to be possible with emotions like enjoyment or delight, although the question arises whether such cases should not be conceptualised as activities rather than states to begin with. In either case, a volitional component appears to be conceivable for such events.

A certain complication for the above distinction of positive and negative emotions arises from Pishwa's characterisation of the contrast between OE *lufian* 'to love' and OE *lician* 'to please', the former of which is described as an "analysable and eventually controllable long-term feeling", while the latter is said to be "an immediate and unanalysed feeling" beyond the control of the experiencer (cf. Pishwa 1999: 133). There are at least two points inherent in these claims. First, the question of whether or not an experiencer is aware of an emotion is taken to impact their ability to exercise control over it. The characterisation of love as an analysable feeling is probably to be interpreted in terms of a heightened cognitive awareness, which results from the assumed salience of this emotion. Evidence of this salience is taken from cross-linguistic instances of nominative case marking of the experiencer argument of verbs denoting love. If salient emotions lead to greater cognitive awareness, this could imply a greater potential for controlling them, and such situations would presumably be more likely to obtain in the case of intensive emotions than in the case of less intensive emotions, which may not even cross the threshold of awareness. This interpretation runs counter to the above observation about negative experiences, however, as being highly salient and thus of overwhelming impact for the experiencer. Cognitive awareness resulting from greater salience of an emotion is thus not necessarily an aspect that implies greater control by the experiencer if it is to be connected with a higher level of intensity of the relevant emotion, and if greater salience is not taken to be exclusive to negative experiences, the distinction between positive and negative emotions likewise becomes irrelevant.

The second point in Pishwa's distinction refers to the durative emotions vs. spontaneous reactions. Her point is taken up by Miura, who considers personal constructions with nominative experiencers a reflection of controllable, long-term feelings, while impersonal constructions are said to encode uncontrollable feelings that arise in the given context of an immediate situation. A potential example of this distinction is the contrast of personal constructions of ME *haten* 'to hate' with variably personal or impersonal constructions of ME

lothen ‘to loathe’ and ME *wlaten* ‘to be disgusted, nauseated’ (cf. Miura 2015: 173). Also Kailuwait observes a cross-linguistic distinction between generalising portrayals of emotion with a subject experiencer and portrayals of episodic emotions with an object experiencer (cf. Kailuwait 2002: 83), although he does not associate this difference with a greater level of control by the experiencer. While durative or generalised emotional states like love or hate may, indeed, present special cases of emotion from a conceptual standpoint, Pishwa’s attribution of a higher degree of control to the experiencer of such emotions does not seem particularly compelling. In the context of impersonal verbs of emotion, the consistent nominative case marking of the experiencer argument of verbs like *love* and *hate* is certainly a notable explanandum, but it does not seem to be accountable for by the assumption of volition or control, unless such properties are circularly inferred from the observed case marking.

The problem of potential circularity is particularly acute in the case of individual instances of deviant case marking. Möhlig-Falke, for example, remarks of a single instance of accusative case marking of the experiencer argument of OE *wundrian* ‘to wonder’ that this may be “motivated by the desire to express a weaker degree of control on the part of the Emoter” (cf. Möhlig-Falke 2012: 160), but she cautions the reader immediately afterwards about placing too much emphasis on this singular occurrence. A more substantial argument seems to be presented for OE *hreowan* ‘to rue’ and OE *sceamian* ‘to be ashamed’, both of which can occur with nominative experiencer arguments in Old English. Möhlig-Falke states that these emotions can be presented as “more or less controlled by the Emoter”, since they may arise from prior contemplation of one’s sins (cf. Möhlig-Falke 2012: 160). As such, it seems that repentance and shame could be conceptualised as cognitive activities rather than involuntary emotional reactions, although a clear separation of the contemplative process and the ensuing emotion would probably still characterise the latter as uncontrolled. A similar argument is presented by Möhlig-Falke for the variable case marking of the experiencer argument of OE *tweogan* ‘to doubt’ and OE *tweonian* ‘id.’. Nominative case marking is attributed to cases where the feeling of doubt arises from contemplation, while oblique case marking is attributed to cases in which the feeling arises spontaneously and is thus arguably less easily controlled (cf. Möhlig-Falke 2012: 161, also Pishwa 1999: 133f.). Since the experience of doubt is probably better conceived of as a cognitive process than as an emotion to begin with, its possible overlap with cognitive activities seems to be a plausible assumption, although a distinction of the two types of situations in grammatically observable terms remains a difficult matter if case marking of the experiencer is not taken as the primary criterion itself.

If emotions at the level of cognitive awareness or emotions that result from a cognitive process can be interpreted as relatively more controlled by the experiencer than emotions that result spontaneously from an immediate impact, this interpretation should be all the more applicable to the experiencer argument of verbs of cognition, which ranks higher than emoters in Van Valin's continuum of thematic roles. A case in point is the personal verb OE *þencan* 'to think', whose semantic range includes cognitive activities like reflection or contemplation, both of which can arguably be controlled at will. The corresponding impersonal verb OE *þyncan* 'to seem', on the other hand, appears to designate a cognitive stance that arises from a given stimulus regardless of the experiencer's volition. Since both *þencan* and *þyncan* are included in the semantic class of verbs of cognition, this means that cognisers are not generally more volitional participants, even if cognitive processes are more likely to be controlled than emotions. An additional distinction between cognitive activities and cognitive states thus appears to be desirable. Such a distinction is, in fact, made by Möhlig-Falke, who applies the latter interpretation as states to OE impersonal verbs like *þyncan* 'to seem' as well as *mætan* 'to dream' and *swefnian* 'id.' (cf. Möhlig-Falke (2012: 93). Also Palander-Collin makes a semantic distinction in terms of different degrees of activity between impersonal *think*, which she characterises as expressing an opinion or a point of view, and personal ("recipient subject") *think*, which is said to primarily express the mental process of thinking (cf. Palander-Collin 1997: 378f.). According to her analysis, the function of the impersonal verb designating an uncontrolled cognitive stance is eventually absorbed by its personal counterpart. The difference between oblique and nominative case marking of the cogniser at later stages is attributed to different degrees of commitment to the expressed opinion, which is thought to be lower in the case of oblique case marking. Her analysis thus implies that not only semantic differences but also pragmatic factors may influence the choice between different types of case marking of the experiencer argument.

The third semantic class of verbs to be considered here are verbs of perception, whose first argument ranks lowest in Van Valin's subdivision of experiencer arguments and thus would be expected to be less identifiable with the role of agent than emoters or cognisers. It seems, however, that, in terms of volition, a similar distinction between perceptive activities and perceptive states like the one for cognition is necessary. The relevant semantic contrast can be illustrated by verbal pairs such as *look* and *see* and *listen* and *hear*, of which only the latter members seem to qualify for a low-ranking perceiver in terms of agency, since they do not presuppose the perceiver's control of the process. The opposite view is expressed by Pishwa, who argues that the presumed activation of short-term memory in the process of perception as

well as the ability to block perception, presumably by closing one's eyes or holding one's ears shut, involve a more active participation than comparable experiences of emotion. It needs to be asked, however, to what extent a language user would be able to recognise or conceptualise a higher degree of mental processing intensity, if this is taken to refer to largely unconscious cognitive mechanisms during the process of perception. In either case, the fact that verbs of perception like *see* and *hear* are personal verbs throughout the history of English does not seem to be related to the degree of perceived volition. For impersonal verbs of physiological perception like OE *þyrstan* 'to be thirsty' and OE *hyngrian* 'to be hungry', on the other hand, a low degree of volition does seem to apply, since the perceptions described by these verbs cannot normally be influenced at will, unless something like active fasting is referred to. It is clear, however, that this interpretation of perception as an uncontrolled process is neither applicable to all verbs of perception (cf. the active verbs *look* and *listen*), nor is it systematically reflected across all verbs where it does apply (cf. the consistent nominative case marking of the perceiver of *see* and *hear*).

Finally, the arguments of verbs that do not express experiences in the narrower sense shall be considered. The first argument of verbs of happenstance seems to lack the property of volition entirely, since such verbs express chance events that occur independently of an optionally profiled experiencer. Verbs of obligation and necessity, on the other hand, may be subject to different interpretations. Both semantic classes express a relational state between two participants, which does not require volitional activity in order to obtain (cf. also Möhlig-Falke 2012: 102–105 and 105–107, who discusses these verbal classes under the headings of "ownership and appropriateness" and "availability and non-availability"). A distinction between internal and external necessity may be relevant in this context, since internal necessity could be conceived of as controlled by the needer in the sense that they are able to define whether or not a given commodity is regarded as necessary (for the assumption of different degrees of volition in the semantic development of OE *behofian* 'to need' cf. Loureiro-Porto 2010: 695). More accurately, the control over what is regarded as necessary seems to lie with the speaker (cf. Möhlig-Falke 2012: 103), who is, of course, identical with the needer in instances of the first person. Even in such cases, however, the participant's control over a situation of need or lack appears to be more limited than in the case of mental, emotional or perceptive experiences, which at least imply some level of internal activity, even though their relative propensity to be conceptualised as active processes is difficult to classify uniformly across all verbs of a given semantic class.

In summary of the present discussion, it seems most reasonable to state that the feature of volition is difficult to identify with the kind of experiencers that are involved in the situations described by impersonal verbs, despite numerous references to more or less volitional or controlled experiences in the literature. A relatively clear distinction appears to apply between experiencers in a narrow sense and participants affected by chance events or relational states expressed by verbs of obligation and necessity. Even so, the kind of internal activity presupposed by the former may not cross the threshold of cognitive awareness and thus not be conceptualised as activity by the language user. The most likely exception to this is presented by cognitive activities like contemplation and, perhaps, semi-active emotional states like repentance, the latter of which may be said to also involve an aspect of cognitive reflection. As such, they can be conceptualised as potentially volitional acts, which may have an impact on variable case marking of the experiencer. In the case of OE *þencan* and *þyncan*, the difference between a cognitive activity and a cognitive state or stance is, of course, lexicalised and only becomes relevant after the eventual merger of these verbs into a single lexeme *think(en)*. More importantly, personal verbs of emotion and perception with consistent nominative case marking of the experiencer illustrate that the construal of these experiences along a continuum of semantic roles and thematic relations is not universally reflected in their construction with either nominative or oblique experiencers. An assessment of the degree of volition which the experiencer argument of a verb belonging to a given semantic class exhibits thus needs to be undertaken on a verb-specific basis. The applicability of volition in a specific context would then have to be argued individually for each instance of the verb, since the differences between more or less volitional involvement are expected to be minute as long as the basic sense of a given verb is preserved.

2.2.2.1.4.2. Sentience/Perception

The second feature in Dowty's list refers to the presence of sensory capabilities in a participant, which is, in fact, the only agent property that Dowty himself ascribes to experiencer arguments (cf. Dowty 1991: 579). For experiencers in the narrow sense, this feature certainly does appear to be a prerequisite, since cognitive, emotional or perceptive experiences require an entity that is capable of experiencing them in order to take place. There is thus little room for variation of this feature with regard to the experiencer argument. The equivalent argument of the semantic classes of verbs of happenstance, obligation and necessity, on the other hand, does not seem to presuppose the feature of sentience, since affectedness by an event or a relation of obligation

or necessity with regard to another participant can also be predicated of inanimate participants without sensory capabilities. Since humans are typically the most relevant participants in the linguistic discourse of other humans, they can be expected to figure more prominently as the first argument of these verbs, while the second argument will typically comprise abstract events, moral judgements or inanimate objects, even though this is not a lexically specified requirement (cf. also Möhlig-Falke 2012: 102 on this distribution with regard to the concepts of ownership and appropriateness). Similarly, the stimulus argument of experiencer verbs in the narrow sense can vary with regard to the presence of sentience/perception, which can be inferred more or less directly from the animacy of a given participant. Animacy thus provides a relatively clear criterion for a distinction between greater or lesser potential for an agentive interpretation of the stimulus argument, although the applicability of such an interpretation in the context of a given predicate meaning remains a separate issue.

2.2.2.1.4.3. Causation

The final item from Dowty's list of agent properties to be considered here is causation. In contrast to volition or control, which have primarily been ascribed to the experiencer argument of impersonal verbs, this feature has repeatedly been attributed to the stimulus argument of such verbs in order to explain the presence of subject coding properties. With regard to nominative case marking, this view is expressed by Anderson (1988: 15), who considers type-ii constructions with a nominative stimulus argument to provide a "full 'causative' interpretation" of the situation compared to instances with a mere "'causative object'" marked by genitive case. Moreover, it is employed to account for the distributional properties of preverbal oblique constituents by Allen (1995: 259), who makes an equivalent distinction between theme and cause arguments. The basic availability of a causative interpretation to stimulus arguments seems to be unproblematic, since the occurrence of an emotion or a cognitive reaction inherently implies some kind of referential target, which can be conceived of as a more or less active trigger of the experience. It seems, though, that an additional distinction between volitional causation and unvolitional causation is, once more, required, since the triggering of a given experience may take place regardless of whether this is intended by the stimulus or whether the stimulus even has the capacity to control the event. Despite the asserted independence of Dowty's agent features, volition thus appears to be a necessary stipulation in order to render causation an appropriate feature for the distinction of more or less agentive stimuli.

The analysis as volitional cause argument is explicitly applied by Allen to the second argument of OE *cweman* ‘to please’ and ME *plesen* ‘id.’, which is contrasted with the supposedly more neutral theme argument of OE *lician* ‘id.’ (cf. Allen 1995: 259, 331). This analysis accords with her distinction of Dative Object verbs, whose subject is consistently assigned to the cause argument, and type-ii verbs, whose subject is variably assigned to either the experiencer or the theme argument. Additional evidence for the distinction is gathered from the distributional properties of these arguments, which involve the optional fronting of the experiencer argument of type-ii verbs in conjunction with a lower-ranking theme, and the frequent realisation of the second argument of the Dative Object verb *cweman* by an animate participant. Since both of these properties are of a statistical rather than of an absolute nature, it seems reasonable to assume that the interpretation as volitional cause can also apply to the second argument of *lician* when it was animate, in which case it exhibited the potential for volitional control of the event. Other verbs of emotion may equally allow for this interpretation if Anderson’s discussion of the constructional patterns of OE *ofhreowan* ‘to cause pity’ is correct (cf. the examples in (35)–(37) above). Interestingly, the nominative constituent *þæs hreoflian mægenleast* ‘the leper’s feebleness’ in (36), which takes part in what Anderson calls the “full ‘causative’ interpretation”, does not exhibit the property of animacy, which would otherwise imply the capability of volitional causation. It is difficult to see, then, how the property of causation can be established for this constituent if not primarily based on its case marking. An animate participant would motivate the interpretation more readily, since the causation of an event can only be volitional if the relevant participant is able to act according to their own will. Clearly, animacy is not a necessary criterion for the assignment of nominative case marking to the stimulus argument of OE *ofhreowan*. It is also clear that animacy does not constitute a sufficient criterion for the actual function of a participant as volitional cause of emotions. This is true for the verbs of liking referred to above, but also, and perhaps even more so, for OE *ofhreowan*, since pity is an emotion that normally applies in relation to another animate participant, even if they take no active part in its causation. Even so, the potential of an interpretation as volitional cause may generally be regarded as higher for animate participants than for inanimate participants.

A causative interpretation is also conceivable for the stimulus argument of verbs of cognition, both in the case of cognitive activities like thinking and cognitive states like surprise. In either situation, the potential for a causative interpretation of the stimulus argument seems to be greater for animate participants. The typical complement of impersonal verbs of cognition like OE *þyncan* ‘to seem’ or ME *semen* ‘id.’, however, takes the form of an internally complex

proposition (e.g. *him seemed that / him seemed X [to be] Y*), which does not exhibit the property of animacy and does not allow for variable interpretations in this respect. The same is true of the second argument of other verbs of cognition like OE *tweogan* ‘to doubt’ and OE *mætan* ‘to dream’. A relation of doubt towards more concrete stimuli like particular persons, abstract concepts or objects would be equally conceivable for *tweogan*, while a more concrete realisation of the second argument of *mætan* would be more likely to indicate the content (or theme) of a given dream rather than its perceived cause. Similarly, the second argument of verbs of happenstance cannot vary in terms of animacy, since the experiencer’s affectedness by a chance event inherently implies a complex situation, which can be either expressed by clausal constituents or by nouns that designate complex events like accidents or misfortune. Since the construal of a participant’s affectedness by such events is optional, it seems that their designation as stimuli is only adequate in the wider sense in which the designation as experiencer is used for the former. The stative situations that obtain between the arguments of verbs of obligation and the arguments of verbs of necessity seem equally unsuited to a causative interpretation of the second argument. The relevant participant is generally not animate in the case of moral obligations, and even if it was predicated of an animate participant that they belonged to another participant, this would treat the animate participant as an object rather than as a volitionally involved cause. It is conceivable, however, that the semantic shift of a verb like ME *wanten* ‘to lack’ from a verb of necessity to a verb expressing something more akin to emotional desire like PDE *want* involves the potential of a more agentive interpretation of the second argument as volitional cause of that desire. Verbs of emotion thus appear to have the greatest potential for a causative interpretation of their second argument.

2.2.2.1.4.4. Summary

The preceding discussion of agent properties has shown that the applicability of such properties to the arguments of impersonal verbs varies considerably across the different semantic classes that can be distinguished. This is not surprising as such, since neither the experiencer argument nor the stimulus argument constitutes a prototypical agent, but both are located towards the middle of the assumed continuum between prototypical agents and patients. Their selection for nominative case marking as a result of their hypothesised identification with the more agentive role is thus a matter of individual semantic properties rather than of a full realisation of the properties of agents. The feature that has been attributed most frequently to the experiencer argument in this context is volition, which can be defined more narrowly as the wilfully

controlled participation in a given event. Among the semantic classes that were considered, this interpretation seemed to be potentially available only to the experiencer argument of experiences in the narrow sense, specifically those denoted by verbs of cognition, emotion and perception, while affectedness by chance events or participation in a relational state of obligation or necessity were regarded as unsuitable for this interpretation. Even in the former case, the experiencer's potential for controlling the situation appears to be generally low, with the possible exception of cognitive activities and emotions that can be conceptualised as consciously maintained and thus as volitional acts. For regular emotions and perceptions, a volitional interpretation of the experiencer seemed difficult to motivate, regardless of their presumed salience or duration. In addition, the grammatical expression of such experiences by personal or impersonal verbs does not follow the simple division of semantic classes in Van Valin's continuum, since emotions like love and hate and perceptions like seeing and hearing are generally expressed by personal verbs with nominative experiencers, despite the fact that this does not seem to be based on a greater affinity with prototypical agents. The remaining cases of oblique-experiencer verbs therefore require an individual assessment of their susceptibility to a volitional interpretation of the experiencer argument.

The applicability of the property of causation to the stimulus argument of impersonal verbs was equally found to be largely restricted to experiences involving cognition, emotion and perception. Since causation was defined more narrowly in the present context as requiring an aspect of volition, its potential implication was found to be most readily inferred from a participant's animacy. While this semantic feature need not immediately imply volitional causation by a given participant, it raises the potential for such an interpretation in comparison to inanimate causes. At the same time, this way of identifying potential causation rules out possible contrasts for situations which involve complex propositional argument as stimulus, since these can be regarded as generally inanimate and thus incapable of volitional causation. This applies to the chance events described by verbs of happenstance as well as the situations in relation to which a cognitive stance is commonly expressed by OE *þyncan*. Animacy is also a prerequisite for the third feature of sentience/perception, which is regularly entailed for experiencer arguments in the narrow sense, but optional for their corresponding stimulus arguments and the participant affected by chance events or participating in a relational state of obligation or necessity, even though human participants presumably constitute the more prominent discourse topic in most of these cases.

2.2.2.2. Hypothesis

Impersonal verbs have played a prominent role in attempts to utilise diachronic change as evidence for synchronic theories of syntax both in transformational paradigm and beyond. Much of the preceding discussion of earlier accounts has thus been devoted to theoretical concepts that are prevalent in this tradition. More specifically, the concepts of reanalysis and actualisation are derived from the notion that the mental representation of language essentially becomes fixed during the process of language acquisition by specification of syntactic rules or variable parameter settings of an innate grammatical system. Under this view, reanalysis takes the form of a more or less instantaneous change involving discrete realisations of a given syntactic feature. The failure of the empirical evidence of impersonal verbs and related phenomena to reflect the relative immediacy of this change can be accounted for by the assumption of additional usage rules, which mediate between the internalised grammar proper and the observable surface manifestations of the parent grammar. The process of elimination of such usage rules in accordance with an innovative analysis has been referred to as actualisation, and this designation appears to be most appropriate in contexts where a preceding reanalysis is actually assumed. Both reanalysis and actualisation have been applied more widely, however, to changes in which an alternative syntactic analysis does not immediately replace but competes with the older one over time, and where the ultimate generalisation of one the variants is regarded as actualisation of diachronic change in a wider sense. This reasoning appears to be applicable to changes in the case marking of oblique-experiencer verbs, even if these are regarded as individual lexical changes rather than the result of a systematic reanalysis of the experiencer argument as nominative subject. The gradual disfavouring of oblique case marking for these constituents, whether it is regarded as a change in grammar-external usage rules or as a change of language proper, conceptualised as a usage based and mutable tool for communication, leaves open the question of how exactly this change progresses through the affected lexical or grammatical contexts.

The notion that diachronic change progresses along grammatically ordered contexts is inspired by Timberlake's discussion of actualisation according to the relative markedness of linguistic environments in relation to a given innovation. Different change scenarios potentially play a role in the complex development of impersonal verbs, including borrowings into Middle English during a phase of renewed productivity of oblique experiencers. If the main process is regarded as an evolutive change, however, the theoretical prediction is that nominative experiencers are found earlier and more consistently in contexts for which the innovative case marking is unmarked. Unfortunately, the theory of markedness, in the form in which it has been

discussed above, does not provide any principled way of identifying unmarked contexts for a given innovation but, instead, suggests that such contexts are specific to a given change under investigation. This entails the problem of potential circularity, since unmarked contexts should not be identified according to the observed progression of a change and afterwards motivated by reference to the concept of markedness. Instead, a well-defined theoretical hypothesis about the contexts in which the innovation is expected to spread earlier and more consistently is required. Second, for the purposes of a quantitative corpus study, these contexts need to be objectively identifiable without the need for extensive qualitative interpretations, which may be potentially subjective and, again, guided by preconceptions about the observed realisation of the feature under investigation. The grammatical distinctions suggested by Timberlake for the identification of more or less volitional or individuated participants are mostly suitable as objectively identifiable criteria, but his identification of agentive nouns in particular appears to rest on individual interpretations of the supplied context, which are unfeasible in the present study. Two points thus require further clarification in order to enable an investigation of the diachronic changes of impersonal verbs, viz. a hypothesis about the contexts in which the innovative nominative case marking is expected to occur earlier and more consistently, and an operationalisation of these contexts in terms of objectively observable criteria.

The central hypothesis to be tested in the present investigation is that the extension of nominative case marking to experiencer arguments of oblique-experiencer verbs occurred earlier and more consistently in contexts with a higher potential for an agentive interpretation of this participant. In quantitative terms, this means that nominative experiencers should occur with a higher proportion in contexts that allow for an agentive interpretation than in contexts that do not allow for such an interpretation. These proportions need to be established within a representative sample of occurrences in a given period of time, since individual occurrences of nominative experiencers in a given environment would not be statistically significant. In functional terms, the hypothesis presents a semantic explanation of the change under investigation, which is contrasted with the traditional notion of a syntax-internal reanalysis of experiencer arguments based on morphological surface ambiguity or constituent-order settings. As such, it has a number of precursors in the literature, although none of these have investigated the hypothesis on the basis of a larger corpus study. A semantic explanation forms the basis of Allen's account of early nominative experiencers with OE *losian* and the use of nominative experiencers in the so-called NO PROP construction of ME *plesen* with an implied propositional stimulus argument, in which the experiencer is the coreferential subject (cf. Allen 1995: 326–328 and 339). In both of these cases, the experiencer is regarded as responsible or

even in control of the event described by the predicate. Conversely, oblique case-marking of the experiencer argument is regarded as an indication of its non-volitional involvement, which is implied by a higher degree of volitional causation by the stimulus argument, to which Allen attributes the semantic role of cause (cf. Allen 1995: 147). A similar argument is presented by Pishwa for nominative experiencers of OE *hreowan*, which she considers to exhibit a higher degree of active participation in their emotion (cf. Pishwa 1999: 133f.). Also Dąbrowska, in her discussion of nominative and dative experiencers in Polish, assumes a general distinction between the foregrounded active role of a nominative experiencer and an essentially passive dative experiencer, which she contrasts in terms of what she calls the craftsman model and the homunculus model (cf. Dąbrowska 1994: 1033). According to Dąbrowska, the prototypical use of the nominative case to express the agent of an action implies a similar construal for nominative experiencer arguments, even if the connection is a less direct one. Similarly, von Steffens-Montag assumes that the frequent correlation of nominative subjects with the semantic role of agent led to the acquisition of the semantic property of volition by the argument of verbs that received innovative nominative case marking (cf. von Steffens-Montag 1983: 196). While the latter statement is made with the opposite implication of cause and compared to the present hypothesis, it still seems possible to assume that the extension of nominative case marking and the attribution of the semantic property of volition to experiencer arguments will be an interdependent process.

According to the above discussion of Dowty's agent properties, the applicability of volition to experiencer arguments needs to be considered extremely limited for verbs that are involved in impersonal constructions in English, with the possible exception of predicates that can be construed as cognitive activities or other kinds of extensions of essentially stative cognitive, emotional or perceptive experiences (cf. also Pishwa's reference to the nominative experiencer argument of *hreowan* as "cognizant"). What seems clear is that any conceivable differences in the degree of volition attributed to the experiencer argument, which essentially need to be based in the language user's ability to detect or conceptualise them, cannot simply be established according to the suggested cline in Van Valin's more fine-grained distinction of experiencer arguments as cognisers, emoters and perceivers, but would need to be motivated independently for any given verb. A reverse argument can be based on the relative potential of the stimulus argument to be identified with the role of agent, however, which seems to imply a relatively low potential of the experiencer to be identified with the role of agent and hence an unvolitional interpretation in comparison to the stimulus. A similar statement is made by Primus about the increased accumulation of patient properties by experiencers when used in

conjunction with so-called control-stimuli, which are identifiable by animacy and the use with volitional *wollen* ‘to want’ in her illustrative examples from German (cf. Primus 2004: 394). Also Möhlig-Falke suggests that the regular syntactic construction of certain impersonal verbs with a nominative stimulus argument in Middle English implies that the experiencer of such verbs “denotes a rather prototypical endpoint or reference point whose agentive properties are minimal” (cf. Möhlig-Falke 2012: 220). Based on these statements, the experiencer’s potential to be identified with the prototypical role of agent and to be selected for nominative case marking on account of the prototypical assignment of this case to agentive participants is to be judged in relation to the respective potential of the stimulus argument, which competes with the experiencer for this identification in a situation of non-prototypical agentivity.

2.2.2.3. Operationalisation

The operationalisation of environments with a higher potential for an agentive interpretation of the experiencer argument of impersonal verbs is a crucial aspect in the design of a quantitative corpus study, not only because a qualitative assessment of individual occurrences based on the presumed implications of their respective contexts would be unfeasible in a study with a large number of observations, but also because such assessments may be prejudiced by the theoretical preconceptions about the distribution of the variants under investigation. Most importantly, the observed case marking itself cannot be taken as an indication of the applicable interpretation of the experiencer, since this would result in a circular argument. In addition, the fact that the investigation is based on historical data implies that native-speaker judgements cannot be adduced in order to disambiguate the applicable interpretation and that no productive tests like the transfer of a given verb into a particular tense-aspect form or the availability of clefting can be applied, since their grammaticality or acceptability would equally have to rely on competence in the given synchronic system. Ideally, the selected criteria should also be informative across all observations of a given verb in order to provide a sufficient basis for their evaluation. While the spontaneous use of a verb in the progressive form, for example, can be taken as indication of a change of state in the experiencer and thus as an additional patient property of this participant (cf. Dowty 1991: 587), the complementary instances of non-use of the progressive are not indicative of either one way or the other and thus do not allow for any conclusions about their observed association with the case marking of the experiencer. Similarly, the conjunction of an observation with volitional ME *willen* ‘to want’ or an embedded adverbial clause of purpose, which would be immediate indicators of a volitional

interpretation, are not contrastive across all instances of a given verb, since the absence of these features cannot be regarded as diagnostic. In view of these requirements, the attraction of grammatical features and categorial distinctions as indicative markers of a more subject-like interpretation like the ones suggested by Timberlake becomes understandable.

There are several limitations to the establishment of potentially relevant distinctions for the development of impersonal verbs in Early English compared to Timberlake's examples from Finnish and Russian, however. First, the distinction between pronominal and nominal realisation that Timberlake makes for subjects of non-finite clauses in Finnish cannot be applied to experiencer arguments in English, since the presumed trigger of reanalysis, the loss of overt case marking on nouns, adjectives and determiners, implies that only pronominal realisations of the experiencer can be readily identified in terms of their construction type, while constructions in which the experiencer is realised by a full NP will typically remain ambiguous during the relevant periods. Second, the distinction between animate and inanimate participants, which is part of Timberlake's hierarchy in the actualisation of genitive subjects in Finnish participial clauses, is not a variable feature in the case of experiencer arguments, at least not for experiences in the narrow sense, since their occurrence presupposes the existence of a sentient participant. This means that no significant contrast is to be expected for experiencer arguments in terms of animacy. Finally, the distinction between agentive and non-agentive nouns, which Timberlake makes with regard to the changes in case assignment to subjects in Finnish infinitival clauses, appears to be based on the author's personal interpretation and not on empirically observable features. Objectifiable criteria for the identification of contexts in which the experiencer argument can be interpreted as more or less agentive thus need to be sought outside the limited variation of eligible experiencer arguments in terms of their categorial realisation and animacy.

A potential criterion that applies to the experiencer argument of impersonal verbs is the more precise definition of its semantic role. The cogniser, perceiver or emoter of experiencer verbs in the narrow sense can be regarded as more agent-like than the so-called experiencer argument of verbs of happenstance, obligation and necessity due to the internal activity required by the relevant experiences. While the extent to which such activity is recognised and conceptualised by the language user cannot be established by the means available to the present study, it seems that no general distinction between more or less agentive experiencers is possible in terms of the semantic distinctions identified in Van Valin's continuum of thematic relations, since both verbs of cognition and verbs of perception involve items referring to either activities or states, even if the former of these are not typically represented by impersonal verbs

in English. Verbs of emotions, for which different degrees of volition of the experiencer have been claimed, appear to permit a more volitional interpretation of this participant when cognitive activity rather than a mere cognitive stance is involved, but the basic experience of emotion is otherwise difficult to interpret as a volitional act. While the expectation deriving from the hypothesis about an experiencer's potential to be identified with the role of agent would thus be that the experiencer arguments receive nominative case marking earlier and more consistently in the case of true experiencer verbs than in the case of verbs of happenstance, obligation and necessity, any additional expectations beyond this distinction need to be formulated individually on the basis of the semantic range of a given verb.

The stimulus argument of impersonal verbs is generally less restricted in its reference than the experiencer argument and thus offers greater potential for variable semantic features that may influence the interpretation of the experiencer as more or less agentive in relation to the corresponding stimulus argument. A distinction between more specific semantic roles similar to the more fine-grained distinction of experiencer arguments has been applied to the stimulus argument of impersonal verbs in terms of a contrast between less agentive theme arguments and more agentive cause arguments, which would imply a correspondingly higher or lower potential of identification with the role of agent for the experiencer. The question remains, however, how the feature of causation can be established independently of the observed case marking or the behavioural syntactic properties of the stimulus argument. One property that was discussed above as implying a higher potential for causation is the animacy of a participant, particularly if causation is defined more narrowly as an intentionally performed act and thus conjoined with the aspect of volition. In fact, animacy, or humanness, implies a higher potential for all of the items in Dowty's list of agent features, including volition, sentience/perception and causation. While it is thus not a very precise indicator of a participant's potential for agentivity, its advantage lies in being more or less objectively identifiable and independent of the contextual interpretation of a participant. Animacy is also one of the features employed by Allen in her distinction of theme and cause arguments of *lician* and *cweman* (cf. Allen 1995: 144–149), and it is interpreted as an agentive property by Möhlig-Falke in her summary of this argument (cf. Möhlig-Falke 2012: 40f.). At the same time, it should be noted that certain relevant predicates from the semantic field of emotion imply the animacy of a stimulus argument regardless of its interpretation as more or less agentive. Specifically, this applies to verbs like OE *ofhreowan* in the sense 'to pity', since this kind of emotion presupposes empathy for another participant, which is only meaningful if this participant is sentient and hence animate (cf. also Wahlén's semantic distinction between

“personal” and “sympathetic” mental affections, which essentially captures this point). In such cases, no or only a very limited degree of variability in the animacy of the stimulus argument would be expected, and no or only a very limited prediction of case assignment to the experiencer would thus be possible.

Two other criteria can be carried over from Timberlake’s discussion of changes in Finnish and Russian, viz. the distinction between pronouns and full NPs and the distinction between concrete and abstract participants. While the latter of these is applied as a criterion in the definition of more or less highly individuated participants in the context of the replacement of genitive case by nominative case in Russian, it seems to be equally applicable to a participant’s potential to be interpreted as agentive, since volitional or causative involvement in a given situation is arguably more readily conceivable for concrete than abstract participants. Abstractness is thus a second semantic criterion according to which a participant’s potential for an identification with the role of agent can be operationalised. There is, of course, a partial overlap in the two features of animacy and abstractness, since animate participants will normally be regarded as concrete, and variation in the second feature of abstractness is thus largely limited to inanimate participants, with the possible exception of the referents of collective nouns, which may be animate but not interpreted as concrete in the same way as individual animate participants. The issue of defining different degrees of animacy and abstractness and the implementation of these two semantic features as predictor variables in a regression model of the diachronic changes in the odds of impersonal and personal constructions will be discussed more thoroughly in the relevant sections on methodology and model building. The second distinction between pronouns and full NPs in Timberlake’s discussion of changes to subjects of non-finite clauses in Finnish seems to be primarily one of categorial realisation and thus of a syntactic rather than a semantic feature, even though the referential properties of a given category may be considered equally as important or even more important for the identification of a participant with the role of agent. While the theoretical basis of the classification of pronouns as the least marked environment for the assignment of morphological subject case to a constituent is not explicitly discussed by Timberlake, his examples suggest that his distinction is primarily directed at personal pronouns or anaphoric pronouns with reference to animate participants. His additional distinction between pronouns and animate nouns seems to imply that the categorial realisation is relevant beyond the shared semantic feature of animacy of the referent. If this is true, then it should be possible to define the levels of a variable representing syntactic categories independently of the semantic features with which they may partly correlate. Categorial realisation can thus be considered as a variable

that is subsumable under the concept of agency according to Timberlake's basic distinction, but, at the same time, it may emerge as an alternative explanatory variable if the necessary distinction of additional levels of syntactic categories cannot be correlated with a respective semantic implication in terms of agency. These issues will equally be discussed in more detail in the sections on methodology and model building.

In summary, it can be stated that the operationalisation of the experiencer's potential for an agentive interpretation is largely based on a reverse argument from semantic and syntactic properties of the stimulus argument. In a situation of non-prototypical agentivity, the experiencer's potential to be identified as the more agent-like participant is raised whenever the respective potential of the stimulus argument is low in comparison. Properties considered as contributing to an agentive interpretation of the stimulus are animacy, concreteness and realisation by a pronominal category. The experiencer's potential to be identified with the role of agent is therefore higher when the stimulus is inanimate, abstract and realised by a non-pronominal category. According to the above hypothesis about the extension of nominative case marking to experiencer arguments of oblique-experiencer verbs as being governed by the semantic interpretation of the experiencer as more or less agentive, this predicts an earlier and more consistent assignment of nominative case marking to experiencer arguments in these contexts. The data base on which this hypothesis will be tested, the method of extracting tokens of the verbs under investigation and selecting observations from the corpora selected for this purpose, as well as the distinction and concrete coding of the levels of the predictor variables will be discussed in the methodological section below, which will also serve to exemplify the use of impersonal verbs in the relevant periods of Middle and Early Modern English in some more detail by means of more or less extensive citations.

3. Methodology

3.1. Corpus Selection

The use of electronic corpora for empirical research into linguistic phenomena has become increasingly prominent in a variety of areas with the increased availability of large-scale open-access resources as well as open-source tools for data management and statistical analysis such as *R* and *CorpusSearch* (for introductions to corpus linguistics in general and to the use of *R* in particular cf. e.g. Friginal/Hardy 2014, Levshina 2015, Gries ²2017). The conditions for corpus-based research into historical stages of most languages are, of course, generally less favourable,

and the available resources are by no means comparable to the variety and vastness of collections of contemporary documents such the *Corpus of Contemporary American English (COCA)* and the *British National Corpus (BNC)*, both of which comprise different genres of written as well as transcripts of spoken texts, amounting to over 560 million and 100 million words respectively at the time of writing. Nevertheless, there are various examples of corpus-based approaches applied to historical data, including several in the domain of impersonal verbs (cf. e.g. Palander-Collin 1996, 1997 on the development of ME *thinken* and Loureiro-Porto 2009 on the semantic predecessors of PDE *need*). One obvious constraint that such studies face is the finite number of preserved texts from earlier periods as well as the dominance of certain genres and the complete absence of spoken records. There may be exceptional cases of new discoveries, but the corpus of extant OE texts is otherwise very circumscribed in both volume and subject matter, comprising mostly narrative texts with ecclesiastic or historical topics, which cannot be modified or expanded to accommodate the needs of a given research question. For Middle English, the situation is, in fact, somewhat more favourable in terms of the sheer quantity of texts, even though these are distributed unevenly across the different ME periods, and also more favourable in terms of the representation of different genres, as additional genres such as personal correspondence becomes available from later ME periods onwards. On the other hand, new problems arise from the particularly high orthographical and dialectal variation, which pose challenges to the still ongoing edition and publication of ME texts (cf. also Markus 1990). The corpus resources available for Middle English at present are thus more restricted than the number of extant documents might lead to expect. This is particularly true for annotated corpora, since their compilation requires a significant amount of financial and human resources, while automated processing is less viable in the case of highly variable texts. As a result, the sample sizes of such corpora are generally limited, which may impact the applicability of inferential techniques on these samples, most notably in the context of low-frequency grammatical phenomena. Diachronic investigations into more recent periods and into changes in progress do not face such problems, as a rule, and they are extremely valuable for our understanding of the general mechanisms of linguistic change (cf. e.g. the early seminal studies by Labov 1966, 1972, Milroy/Milroy 1978). For specific historical changes as well as long-term diachronic developments, reference to the restricted resources outlined above is unavoidable, however, and they should not be dismissed despite their limitations.

A valuable tool for surveying the available resources is the searchable *Corpus Resource Database (CoRD)* hosted by the University of Helsinki, which presents a list of both synchronic and diachronic corpora of English (for an overview of the latter, cf. also Rissanen 2000,

Friginal/Hardy 2014: 180–187, Möhlig-Falke 2016: 395–397). The most prominent long-term diachronic corpus is probably the diachronic part of the *Helsinki Corpus of English Texts (HC)*, which comprises texts written roughly between the 8th century and 1700 AD. These amount to a total of just over 1.5 million words, which makes the corpus rather small compared to its contemporary counterparts, particularly if the periods are considered individually (Old English = 413,250 words, Middle English = 608,570 words, Early Modern English = 551,000 words). Another point to consider is that neither the original version nor the more recent TEI XML release of 2011 provides any linguistic parsing of the texts, which are presented in their raw form alongside codes for a number of general parameters such as known authors, date of composition and genre, as well as bibliographical information on the editions underlying the texts. Annotated and significantly expanded versions of the various parts are available in form of the separate *Brooklyn-Geneva-Amsterdam-Helsinki Parsed Corpus of Old English*, which has been superseded, however, by the *York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE)* and the *York-Helsinki Parsed Corpus of Old English Poetry (YPC)*, as well as the *Penn-Helsinki Parsed Corpora of Historical English*, which include several editions of the *Penn-Helsinki Parsed Corpus of Middle English (PPCME2)* and the *Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME)*. All of these use versions of the Penn-Treebank annotation system for part-of-speech (POS) tagging and syntactic parsing (PSD) in the form of bracketing and tags (cf. Taylor et al. 2003). The smallest of these corpora is the one devoted exclusively to OE poetry, which contains just over 0.07 million words, while the other corpora are comparable in size to the original *HC* with around 1.5 million words (*YCOE*), 1.2 million words (*PPCME2*) and 1.8 million words (*PPCEME*) respectively. Apart from these derivatives of the *HC*, there are the independent *Parsed Corpus of Early English Correspondence (PCEEC)*, the *Corpus of Middle English Prose and Verse (CMEPV)*, which is part of the University of Michigan's *Middle English Compendium*, and the *Innsbruck Computer Archive of Machine-Readable English Texts (ICAMET)*, which includes the *Innsbruck Corpus of Middle English Prose (ICoMEP)* and the *Letter Corpus of ICAMET*. A notable feature of the latter two corpora is that they comprise only complete versions of texts, which makes them especially suited to discourse-related analyses. The prose corpus is also particularly large with around 7.8 million words, while the *Letter Corpus* is relatively small with only 182,000 words. Neither corpus includes part-of-speech tagging or syntactic parsing, however, which means that they cannot be queried as readily for grammatical features. This is also true for the *CMEPV*, which comprises almost 300 texts of varying length since its most recent revision in 2018 but can only be searched for simple or combinations of character strings. The *PCEEC*, on the other hand,

contains both levels of syntactic analysis in the same format as the other parsed corpora described above. It is part of a family of corpora, which includes an unpublished original version, a significantly reduced sampler, and an extension and supplement to the original version, which still await completion at the time of writing. The parsed version covers the period between 1410–1681 and contains just under 2.16 million words, which makes it somewhat smaller than the original unpublished version with close to 2.6 million words, although it still constitutes a significant body of annotated texts. Finally, although not a corpus in the strict sense, the *Dictionary of Old English Corpus (DOEC)* is a noteworthy resource, since it contains at least one version of every extant OE text alongside a certain number of parallel versions, which comprise a total of over 3 million native and around 0.75 million foreign words. This makes it a near-exhaustive treatment of the period, but syntactic parsing and philological information is entirely absent from the collection. Other diachronic or historical corpora are either too late in their coverage (e.g. *A Representative Corpus of Historical English Registers (ARCHER)*, with documents from 1650–present) or too specialised (e.g. the *Middle English Medical Texts (MEMT)*) to be considered in the present context.

While several corpora are thus available for research into earlier periods of English, none of them appear to be without individual drawbacks for the investigation of impersonal verbs (cf. also Miura 2015: 96f.). The individual sections of the diachronic part of the *HC* are too small to serve as a basis for an investigation that involves a low-frequency phenomenon or any larger number of variables, and its primary value thus lies in the initial exploration of long-term diachronic data and the conduction of pilot studies. Some researchers have also chosen to expand the corpus with other materials (cf. e.g. Palander-Collin 1997: 375f., Loureiro-Porto 2009: 11–14) to arrive at a more balanced and more solid database. The *DOEC*, on the other hand, is a near-exhaustive resource for Old English, but it lacks the syntactic parsing provided only by the smaller collections of the *YCOE* and the *YPC*, which combine to only about 52% of the *DOEC*'s word count of native words (cf. Möhlig-Falke 2016: 396; note, however, that the *DOEC* also includes almost 0.7 million words from interlinear glosses, which means that the parsed corpora's coverage of prose and poetry is, in fact, closer to 67%). For Middle English, no resource with comparable coverage of the admittedly more numerous extant sources is available, even though the *ICAMET* with 7.8 million words and the *CMEPV* with almost 300 full texts, whose precise word count is not specified, are comparatively large. Miura (2015: 96f. with references) advises against the use of the former corpus for an investigation of impersonal verbs, however, since it is entirely restricted to works in prose, while the occasional use of verbs in impersonal constructions occurred most prominently in poetry in her investigation. Given

the importance of poetry within the ME canon, this may be a reasonable objection, and as noted above, the *ICAMET* was designed with a specific aim, namely that of discourse analysis, in mind. Nevertheless, the unusual occurrences of verbs in impersonal constructions may simply be more frequent in poetry due to poetic licence, which would mean that their exclusion would not necessarily reduce the representativeness of a prose corpus of actual spoken language in this respect. The lack of syntactic annotation imposes serious limitations on the usefulness of the larger collections, however, since the retrieval of specific forms from raw text files cannot be based on any grammatical criteria. The *PPCME2* and *PPCEME*, which make up for this lack, are much smaller in comparison and share the *ICAMET*'s bias towards prose, while the *PCEEC* comes with the additional problem of being restricted to a particular genre.

The ideal resource for an item-based investigation like that of impersonal verbs would be a lemmatised corpus with the immediate option of retrieving all instances of a given lexeme within the corpus. This is especially true in view of the many different variant forms that lexical items can take during the ME period, which makes the definition of a finite set of forms that a query should include much more difficult. Three major factors contribute to this variability: 1. the greater stock of morphological forms, arising from both stem formation and inflection, 2. the variation between regional dialects, 3. the extreme orthographic variation in the representation of these forms arising from the lack of an inter-regional standard (for overviews of the principal verbal endings in the major dialect areas cf. Mossé 1952: 76, 80, Fisiak 1968: 92, Markus 1990: 156f.). The ME ending of the third-person singular present indicative, for example, varied between forms ending in /z/ (typical in the North) and forms ending in /θ/ (the traditional form in the South and Midlands), which could be preceded by a vowel most commonly spelled <e> or <i>, although also <u> and <y> or the absence of a vowel were possible, and for which <s, ʒ, z> and <þ, ð, th>, either simple, geminate or in combination with each other (cf. Fisiak 1968: 22), were the most common graphemic variants besides <t> and <d> or even <y> for the latter. While this amount of variation may appear structured and manageable, it only covers the major variants of a single morphological category (other morphological categories with dialectal variants worth noting include the present participle in *-and(e)*, *-ind(e)* and *-ing(e)* and the present indicative plural in *-es*, *-en* or *-ep* together with their respective graphemic variants), and the variation covered by grammatical treatments generally does not exhaust the actual variation encountered in ME manuscripts. The latter is represented much more closely in the *Linguistic Atlas of Late Mediaeval English (LALME)*, which records variants such as *-ce -eʒth*, *-oth*, and *-ysse* (Northern) as well as *-at*, *-eht*, *-yghtʒ* and *-yGihthe* (Southern) besides many others for the third-person singular present indicative.

Such variation clearly poses a challenge to item-based investigations of earlier stages of English and contrasts sharply with the relatively limited set of verbal forms that need to be considered for the grammatically and orthographically standardised varieties of Present-Day English.

None of the corpora discussed above are published in lemmatised form, however, and some of them do not even include POS tagging or syntactic parsing, which would enable searches for specific syntactic categories or constituent structures (for supplemental tools to the *PPCME2* cf. Percillier 2016 and the discussion in section 3.2 below). The *DOEC* comes with a tool that returns lists of potential orthographic variants of a search item (cf. Möhlig-Falke 2012: 26, 2016: 399), and there are other licensed tools available for the retrieval of variants based on pre-defined substitutions. Such tools definitely improve the chances of finding forms of a given lexical item, if all morphological and dialectal variants are anticipated correctly, but the assignment of the resulting orthographic forms to a specific lemma still requires manual labour. The intensity of this task is aptly demonstrated by the long lists of “forms scrutinized” in the appendices to Loureiro-Porto’s study (cf. Loureiro-Porto 2009: 220–249), which, in many cases, did not yield instances of the desired verb (for example, only 40 out of the total 6,820 forms investigated in the *HC* were actually forms of ME *thurven* and *bethurven*). Given these complications, the use of an already lemmatised resource such as the *Bosworth-Toller Anglo Saxon Dictionary (BT)* or the *Middle English Dictionary (MED)* becomes an attractive alternative. While citations in the slightly dated *BT* and its addenda are probably too few to function as the basis of any real syntactic analysis, the *MED* seems to be a more viable candidate. As part of the *Middle English Compendium*, it certainly is the most comprehensive, albeit not exhaustive, treatment of Middle English to date, and owing to its exemplary citations of all recorded uses of a lemma throughout the covered period of 1175–1500, it forms a sizeable corpus of its own. The *MED* is, in fact, used as the primary database by Miura, who names its size, lemmatisation and the inclusion of low-frequency items among the advantages for her study of a particular lexical field of impersonal verbs (cf. Miura 2015: 97f.). Her design is explicitly qualitative, however, since the *MED*’s practice of supplying “at least one quotation” of each sense for every “quarter century” (cf. Lewis/Williams ²2007: 18) gives some indication of the overall frequency of a verb in a particular sense or construction, but it provides no solid basis for a quantitative study with the aim of deriving frequency-based statistical inferences (cf. Miura 2015: 63, 97f.). Another drawback is the limited context provided for each quotation, which is not always sufficient to judge the syntactic and semantic properties of the verbal arguments involved. Despite their limitations in terms of user-friendly processing and despite remaining shortcomings in terms of balance according to genre and period, some of which are

obviously pre-determined by the circumstances of ME transmission, the use of actual corpora rather than lemmatised dictionaries thus remains a requirement for any study aiming at a statistical analysis of the data.

For the present study, the *PPCME2* and the *PCEEC* were selected as the basis of the analysis. The primary reason for this is that they are relatively large according to the applicable standards of historical corpora (1.2 and 2.16 million words respectively) and thus lend themselves more readily to the investigation of syntactic structures with a relatively low frequency than the significantly smaller sections of the original *HC*. Second, these corpora are parsed and thus facilitate searches that incorporate grammatical information, which contrasts with larger corpora like the *ICAMET*, whose use would otherwise have been desirable. Third, they combine to span roughly the period in which changes in the argument structure of impersonal verbs are most likely to be detected (1150–1681). The *PPCEME* would have provided a viable alternative to the *PCEEC*, although it is slightly smaller than the latter (close to 1.8 million words) and also slightly later, since it lacks the additional ME data contained in the latter. In either case, the expansion of the *PPCME2* with further materials was necessary in order to arrive at a more substantial database, and this practice of combining several sources follows the aforementioned studies by Palander-Collin (1996, 1997) and Loureiro-Porto (2009). A combination of data from the *PPCME2* with quotations from the *MED* was also considered but ultimately rejected, since it would have skewed the frequency distributions and required the use of additional sources in order to determine the context of the *MED* quotations where necessary. Using the *PCEEC* as an additional resource obviously comes at the cost of genre balance, since all data after the ME period will exclusively reflect a single genre, viz. personal communication. Admittedly, the *PPCEME* would have been preferable in this respect, since it covers a variety of genres such as diary entries, drama, medicinal tracts and travelogues. It also displays greater balance in terms of periods, as its contents are distributed relatively evenly across the three EModE periods that are customary in the Helsinki tradition (E1 (1500–1569) = 576,195 words, E2 (1570–1639) = 652,799 words, E3 (1640–1710) = 565,016 words), while the *PCEEC* has a pronounced bias towards E2, which constitutes over 40% of the entire corpus. On the other hand, the specific genre represented by the *PCEEC* holds the potential advantage of reflecting spoken language more closely than most other written genres, and it may thus be a more sensitive diagnostic of linguistic change in progress than the general corpora (cf. Raumolin-Brunberg 1996: 16f., Nevalainen/Raumolin-Brunberg 1996: 39). This expectation seems to be borne out by the earlier and more rapid generalisation of the evidential use of personal *thinken* in the *PCEEC* compared to the *HC* observed by Palander-Collin (1996: 140).

Some qualification is necessary, however, since the letters that have come down to us from earlier periods of English obviously do not represent all strata of society equally well but are naturally limited to those who had access to literacy, which includes predominantly male members of merchant families and the aristocracy. These facts lie beyond the influence of the compilers of the corpus, of course, but it is worth noting that the correspondence corpus, despite its merits over corpora with predominantly literary genres, is, perhaps, not as closely representative of the spoken language of the time as initial intuition might suggest.

In combination, the *PPCME2* and the *PCEEC* comprise just over 3.3 million words, whose distribution across the ME and EModE subperiods based on manuscript and letter dates is detailed in table 1 below:

	PPCME2	PCEEC	total (%)
M1 (1150–1250)	258,090	NA	258,090 (7.8%)
M2 (1250–1350)	93,999	NA	93,999 (2.8%)
M3 (1350–1420)	403,007	19,505	422,512 (12.8%)
M4 (1420–1500)	400,869	364,317	765,186 (23.1%)
E1 (1500–1570)	NA	309,220	309,220 (9.3%)
E2 (1570–1640)	NA	910,675	910,675 (27.5%)
E3 (1640–1710)	NA	555,415	555,415 (16.8%)
total	1,155,965	2,159,132	3,315,097 (100%)

As can be seen, the most poorly represented period is M2, which contains only about 2.8% of the total number of words, while the most highly represented period is E2 comprises more than one fourth or about 27.5% of all words, which clearly exceeds the 14.3% expected in a perfectly even distribution. This unevenness is already inherent in each separate corpus and reflects the relative productivity and the circumstances of transmission of each period, which is not simply a matter of chronological distance, since M1 contains more words (c. 7.8%) than M2, and E3 contains less words (c. 16.8%) than E2. Taken together, however, the two earliest periods M1 and M2 comprise 10.6% of all words, while the two latest periods E2 and E3 comprise 44.2% of all words, which presents a clear bias towards the later periods. While the inclusion of the later periods is desirable in principle, this heavy bias is not necessarily favourable for the investigation of impersonal verbs given that their development is supposed to have been largely

concluded by the end of the 16th century (cf. van der Gaaf 1904: 142). In addition, the central period M4 benefits disproportionately from the overlap between the two corpora, providing 23.1% of the total number of words combined.

In terms of genre, the *PPCME2* constitutes a relatively balanced general corpus, although there is a certain preponderance of ecclesiastic writings, particularly in the earliest period M1, which is due to the dominant production of such writings at that time. This means that the corpus may be representative of the period's literary activity, but access to genres that are more closely related to oral modes remains limited. The compilation of the *PPCME2* comprises 55 text samples, which are largely expansions of the samples contained in the *HC* with only minor additions and deletions. They include religious treatises, homilies and rules, but also histories such as the *Peterborough Chronicle* and *Layamon's Brut* as well as functional texts such as laws, documents and handbooks as well as literary genres such as romance, fiction and, in the latest period M4, drama. The *PCEEC*, on the other hand, which was initiated as a resource for historical sociolinguistics (cf. Nevalainen 1996: 3 and the pilot studies in the same volume), consists exclusively of samples of private letters, which are taken from over 80 previously published editions. The collections range in size between several hundreds of words to over 200,000 words (the *Paston Letters*), although the majority contain between several thousand and several tens of thousands of words. While the letters are generally autographs, they are not necessarily all of a personal nature but include official letters as well (cf. Nevalainen/Raumolin-Brunberg 1996: 54¹). In terms of regional provenance, they are centred around the areas of London, Norfolk and the North, but they are otherwise intended to reflect demographic variants such as social status, mobility, age and gender as equally as possible. This aim cannot, of course, be fully realised for the non-gentry and women in earlier periods, since these groups often lacked the literacy required for active participation in written correspondence (cf. Nevalainen/Raumolin-Brunberg 1996: 41f.). While this shortcoming may affect the overall representativeness of the corpus, the demographic markers recorded in the *PCEEC* were not utilised for the present study and do not constitute a particular point of research. It should merely be noted as a general caveat that the collections predominantly reflect the linguistic conventions of the educated, male gentry and members of the professions.

In summary, it is clear that a great amount of work has been invested into the development of corpus resources in order to facilitate synchronic and diachronic research into earlier periods of English, for which unreserved credit is due to the compilers. The extension of current research methods to historical data constitutes a significant advancement of the field, since it greatly enhances the scope and the ease with which large amounts of data can be utilised

for hypothesis building and testing. Nevertheless, a number of limitations remain, especially when compared to the vast resources available for contemporary varieties of English. Some of these are dictated by transmission. The amount of extant historical data is finite, and the availability and representation of different text types, regional varieties and demographic variables, particularly during the earliest periods, is uneven at best. For the later ME periods, however, the written sources are more extensive and simply not exhaustively covered in the existing corpora, which means that the continued editing and compilation of these sources are a desirable future endeavour. An additional problem in this context is the high degree of morphological, dialectal, and orthographic variability of the sources, which requires automated parsing methods to be complemented by a significant amount of cost-intensive manual labour. This results in smaller corpus sizes for the syntactically annotated corpora compared to the existing larger collections of plain text. Syntactic annotation is a requirement, however, if corpus queries are supposed to include grammatical information, and, as such, annotations are indispensable for many research questions. Investigations of low-frequency phenomena require greater sample sizes of pre-processed data, and the absence of lemmatisation in all published resources except the dictionaries is a particularly disadvantageous shortcoming for research into specific lexical items. Of course, individual solutions to these problems are possible (cf. Percillier 2016 and below), but an expansion of the currently existing resources for general access in the areas outlined above would still be desirable. As it stands, the present study will need to proceed with the resources available at the time of composition. The combination of the *PPCME2* and the *PCEEC* as the two major resources results in a substantial database that seems reasonably suited to the investigation of impersonal verbs despite the remaining compromises with regard to chronological distribution and genre representation. The process of extracting the relevant data from these corpora will be detailed in section 3.2 below.

3.2. Data Extraction

As mentioned above, none of the official releases of the corpora used in this study natively include information about the association of their items with particular lemmata. For the *PPCME2*, an independently released lemmatiser is available, however, both in the form of a downloadable script and as a web-based application, which was developed as part of the *Borrowing of Argument Structure in Contact Situations (BASICS)* project conducted at the universities of Stuttgart and Mannheim (cf. Percillier 2016: 209 and 210–212). Since the methodological aim of that project is comparable to that of the present study, its procedure will

be described in some detail. The first step in acquiring the lemma information of a given verb involved the manual assignment of graphemic verb forms to lemma-specific IDs or to multiple IDs in cases of doubtful assignment. The IDs themselves were derived from the relevant *MED* entries. A total of 19,320 forms were paired in this way with 2,979 primary and 4,973 other possible matches. The next step involved the insertion of lemma information into the corpus based on this inventory of correspondences. This was achieved either directly by assigning a given form to its counterpart ID wherever possible, or, if no direct match was achieved, by recourse to one of two modifications, viz. the creation of graphemic variants by means of a set of pre-defined substitution rules (e.g. *i > e / y*), or by stemming through the attempted removal of a specified set of ME inflections and their comparison to stemmed forms in the inventory. According to the report by Percillier, this procedure performed well and led to 110,116 immediate assignments out of 130,282 total verb forms in the corpus (c. 84.5%). Another 5,868 forms (4.5%) and 10,421 forms (8%) were assigned by graphemic substitution and stemming with a sample-based estimated accuracy of 86% and 90% respectively, which left only 3,877 forms (2.98%) unassigned. Despite the good performance, however, a couple of open questions still remain. Since the graphemic substitutions and the stemming procedure were based on pre-defined sets of substitution rules and inflectional endings, less typical forms are presumably excluded from the lemmatisation. This is, of course, a necessary trade-off and appears to have had only limited impact given the low number of unassigned forms, which, nevertheless, require additional manual coding if forms representing sub-standard conventions are to be included. Also, the extent of ambiguous assignments is not immediately apparent from the above figures, since alternative matches defined during manual coding are presumably distributed unevenly among the forms. While many of the primary matches may be clear-cut, others will have multiple alternative options. On average, though, each primary match is associated with at least one alternative match, and this rate of ambiguity is necessarily transferred to those forms that were assigned to their lemmata based on the initial inventory. This is not so much a criticism as an observation, since the presence of potential homonyms and homographs is inescapable and ultimately requires the identification of such items by their context in any case. An option of calling forms with a high probability of belonging to a certain or several potential lemmata already presents a significant improvement over the unlemmatised corpus. The availability of the *BASICS* research tool was noted too late, however, for utilisation in the present study. Given that the resources available to the present study for manual processing did not match those of the *BASICS* project (about 17.5% of the immediate assignments, representing 14.8% of all verb forms in the corpus, were performed manually by a team of research assistants in the *BASICS*

project) and given the narrower scope of this investigation, the method selected for the present purposes was more compromising and more pragmatic by comparison. Since the inclusion of an additional corpus besides the *PPCME2* was envisaged, an independent method would have had to be developed for queries in that corpus in any case, and, even for the *PPCME2* itself, the resulting numbers of identified forms appear to be quite comparable to those identified by the *BASICS* lemmatiser. For example, the lemma search of the *BASICS* web application returns 10 types and 51 tokens of ME *liken* ‘to please’ (MED25553), most of which are also referenced as potential forms of *liken* ‘to compare’ (MED25554). In the present study, 257 potential forms of *liken* were identified initially, and 188 of these were ultimately classified as certain instances of *liken* ‘to please’ (MED25553). The latter figure is comparable to the total of 194 forms cited by Trips/Stein (2019: 249) in their work derived from the *BASICS* project. The process of identifying these forms in the present context will be described in the following paragraphs.

The *PPCME2* and the *PCEEC* are distributed in three different formats, viz. text files, part-of-speech tagged files and syntactically parsed files. The text samples in all of these formats are divided into individual tokens, which are marked by a unique identifier and usually comprise a main clause and its subordinate clauses. Part-of-speech tags and all other types of coding in the tagged files are separated from the elements they refer to by an underscore, while syntactic coding in the parsed files is performed by labelled brackets, which also include the POS information of the individual items. Since the goal for this study was to extract all instances of specific lexical items irrespective of their complementation patterns, recourse to the syntactically parsed files held no additional advantage. Although syntactic tags for empty categories including empty expletive subjects (coded *exp*) are available in the corpus and could be queried with the help of a dedicated programme like *CorpusSearch*, such constituents are only expected to occur in a small subset of impersonal constructions in which neither an experiencer argument nor a stimulus argument nor an expletive performs the role of subject. Conversely, empty expletive subjects are not limited to instances of impersonal verbs. Queries were therefore executed in a *Python* script instead, which utilises the POS information included in the tagged files in order to restrict the results to lexical verbs. The second aspect was the specification of potential graphemic forms of the verbs to be identified by the queries. This task was realised in a series of steps.

The first step consisted in the creation of a basic inventory of forms. This was done with the help of the *MED*, which typically exemplifies the different uses and senses of a given headword by more or less extensive quotations. The *MED* entry of *liken* ‘to please’, for example, contains 257 quotations (and 1 additional quotation in the supplementary material,

which was not utilised). All of these quotations contain at least one token of the verb and, in some cases, additional variant readings, the latter of which were not utilised in the case of *liken*, although their inclusion would have been preferable in order to ensure the representation of as broad a variety of forms as possible, and all subsequent queries hence took such variants into account. Based on these tokens, 60 different types of morphological and graphemic forms were established (e.g. *liketh*, *likeþ*, *lykede* etc.). In the next step, each type was separated into its graphemic units. The basic criterion for this was the representation of phonemes contained in the verbal base and the desinence, which, in the case of *liken*, include at least the three segments of the root, viz. the initial radical /l/, the root vowel /i:/ and the final radical /k/, and up to four additional segments in the desinence (in the case of preterit stems). The task may seem trivial at first, but not all divisions were straightforward, and some subjectivity may be involved. For example, the root vowel of *liken* was spelled in four different ways in the *MED* quotations, viz. <e, i, ij, y>. Since <ij> was interpreted as a marker of vowel length (or a diphthong), <j> was not recorded as a separate segment initially, but as part of a digraph corresponding to the simple spellings. Similarly, character combinations, geminate spellings and simple spellings were treated as equivalents in the third-person singular present indicative ending containing the interdental fricative usually spelled <ð, þ, þþ, th> in the quotations (additional spellings of this form with <t, d> are only distinguishable from spellings of the preterit marker by grammatical analysis, which, in the present context, can be ignored). The same principal was applied to the vocalic component of the desinence, which is most commonly spelled <i, y, e, æ, a> in the quotations, but also <ie, iæ, ia>. This presumably continues the more complex stem formation of OE *lician* < Proto-Germanic *-ō-*īe/o-*, which means that the two vowels in fact represent two different phonological segments, which are only equivalent to the simple spellings as a morphological unit. The phonologically accurate analysis of <ij> as a single unit, on the other hand, was later revised by extrapolation of <j> (and also <y>) as a potential marker of vowel length (or diphthongal articulation) to all characters representing the root vowel (<i, y, e>). In this way, it was possible to query all possible combinations of the attested spellings and ensure that no combination of a particular spelling of the stem with a particular spelling of the endings would be missed simply because it was not attested in the quotations. For example, the form *lykide* is present in the *PPCME2*, but *lyk-* and *-ide* are only attested individually in the quotations and not in combination. The graphemic units established in the above process were combined in a script by regular expressions, which state the optionality or obligatoriness of each set. For example, the representation of the first radical of *liken* was specified as obligatory <l>, and the representation of the root vowel was specified as an obligatory choice among <e,

i, ij, y>. All character sets after the root are essentially optional, since a string such as <lyk> can already represent a complete form. The first character of the desinence was thus initially specified as an optional choice between <a, æ, e, i, ia, iæ, ie, o, y>, <æ> and <iæ> being represented by \+a and \+a in the script in order to accommodate the *PPCME2*'s and *PCEEC*'s convention of transcribing non-ASCII characters (similarly $\delta = +d$, $p = +t$, $z = +g$, always preceded by the escape character \ in the script). The character <o> was derived from the preterit and the participial stem, cf. *licode*, *likonde* in the *MED* quotations, but no such form was encountered in the corpora.

The initial results of this approach were moderate. The search with graphemic units combined by regular expressions yielded 189 hits in the *PPCME2*, which constitutes an increase of 22 hits (c. 11.6%) compared to the 167 hits generated by the simple item-based search with verbal forms from the *MED* quotations. Note, however, that these numbers merely represent potential forms of the verb, which require further manual verification in either case. The POS filter implemented in the script already eliminates a significant number of false-positive hits, including adjectival forms, adverbial forms and foreign words, which account for 171 or over 50% of a total of 338 hits in an item-based search without specification of the part of speech (false-positive hits of nouns are generally ruled out at this point due to the fact that the desinence was specified based on verbal forms, but cf. below). This illustrates the usefulness of POS tagging for this type of search, especially when derivations of a queried item in other syntactic categories are to be expected. Several attempts at improving the result were conducted, including the aforementioned generalisation of the digraphic representation of the root vowel by means of <j> (and additionally by <y>), the addition of <cc> and <ck> as possible representations of the second radical, which was already represented by the geminate spelling <kk> in the *MED* quotations, as well as the addition of <n> (and <nn>) to the character set in fourth position after the root, which, despite its absence from the quotations, is a character potentially expected in the third-person plural desinence *-ede(n)* (cf. e.g. Fisiak 1968: 92). None of these modifications had a very pronounced impact, however. The only preterit form including a final nasal elicited from the *PPCME2* was *lykkyden*, and this proved to be a false-positive form of *likken* 'to lick', while the *PCEEC* provided no such forms, presumably because the phonological reduction of the desinence was already completed by the time of the earliest documents included in that corpus. Digraphic spellings of the root vowel including <j> or <y> did not occur at all in the *PPCME2*, while the *PCEEC* contained two forms *leyck* and *leyke*, both of which were verified as forms of *liken* 'to please'. The addition of <ck> yielded 7 additional forms in the *PPCME2* including *licke*, *lycked* and *lyckyn*, but all of these belonged

either to *likken* ‘to lick’ or to *liken* ‘to compare’ (MED25554). This fact demonstrates that the increased scope of a search may lead to a decrease in accuracy, which needs to be considered when setting the search parameters. It also demonstrates that a certain degree of inaccuracy is unavoidable, since a spelling that represents *liken* ‘to compare’ can, in principle, also be expected for *liken* ‘to please’, and, indeed, the *PCEEC* contains several forms with a <ck> spelling that represent *liken* ‘to please’ (cf. *leyck*, *licke*, *lickes*, *lycke*) besides forms that represent *likken* ‘to lick’ (cf. *licked*, *lickt*). Overall, the results of the above modifications show that the query with graphemic units combined by regular expressions is able to retrieve a considerable number of potential instances of the desired lexical verb when combined with the POS information of the tagged corpus files, but since it is based on the specific inventory of forms present in the MED quotations, which, although numerous, are not specifically geared towards representativeness of graphemic variants, it is unlikely to provide exhaustive coverage of all forms that are to be expected.

In view of the apparent restrictions of the basic inventory of forms, further attempts at improving the query were made by reference to the *Linguistic Atlas of Late Mediaeval English* as an additional resource for attested desinences. As mentioned above, the atlas includes lists of diagnostic dialectal variants, which also comprise certain grammatical items like the third-person singular present indicative ending. The coverage of this form in the *LALME* is, by design, much greater than that of a single entry in the *MED*, which is easily apparent from the fact that most forms containing some of the more unusual endings like *-e3th* or *-yght3* do not occur in the entry of *liken* ‘to please’. A search utilising the third-person present indicative and the preterit endings recorded in the *LALME* resulted in 213 hits in the *PPCME2*, but although some of these revealed additional endings such as *-up* and *-yp*, which happened to be not recorded in the *MED* entry, the majority of the less common endings did not occur in the corpus. In view of the low chances of encountering these endings, their explicit specification appeared less desirable in the context of the query, since the inclusion of additional characters simultaneously increases the chances of generating false-positive hits. In addition, not all inflectional endings are included as diagnostic forms in the *LALME*, a fact which may lead to an unwanted bias. While verbal forms in impersonal type-i constructions will be restricted to the third person, nominative stimuli and the spread of nominative experiencers entail that such verbs are open to use in all grammatical persons, which means that these need to be represented in the query. Of course, the endings of the first-person singular present indicative and the singular subjunctive are phonologically less complex than, say, that of the third-person singular present indicative, which means that they will be already covered by the more complex endings,

since these are segmented into optional graphemic units in preparation of the query. If the preservation of inflectional endings as discrete units is not a stipulation of the query to begin with, however, the prediction of all possible endings and their graphemic variants becomes superfluous, since this task can, effectively, be delegated to the integrated POS filter, which automatically selects all possible verbal forms occurring in the corpora based on the pre-established POS tags. The main task, then, is to define possible variants of the lexical core, viz. the verbal root, in order to restrict the results to forms of the desired item as much as possible. Since the verbal root is present in all instances of a given verb, the chances of obtaining a representative sample of graphemic variants from the *MED* quotations is relatively high, even though modifications of the sample may still improve the overall results. All positions after the root were thus ultimately specified as optional sets of all possible characters in the final version of the query. This ensures complete coverage of all verbal forms that contain one of the possible character strings specified for the root, regardless of whether they contain a more common or a less common desinence, and regardless of the grammatical category they represent.

For obvious reasons, the relaxed precision of the query generates a certain number of false-positive hits. This includes verbs with more complex stem formations, e.g. *licensen* ‘to license’, which are usually conspicuous among the search results on account of their length and can be sorted out easily upon manual inspection. A slightly more difficult case is presented by hits belonging to verbs such as *lechen* ‘to cure’, *lighten* ‘to lighten (a load); to descend’ (cf. also *lighten* ‘to light’) and *liknen* ‘to compare’. These arise, in part, from a combination of the final character of the root with one of the subsequent characters, resulting in a digraph that can be used to represent roots other than that of *liknen*. Apart from this, *lighten* and *liknen* each contain an additional segment that makes the verbal base more complex but does not render it as easily detectable as that of *licensen*, since some of the forms actually attested in the *PPCME2* could be considered possible spellings of *liknen*, cf. e.g. *leche* ‘to cure’, *licht* ‘shines’, *likne* ‘to compare’. Although some of these could be eliminated by restricting the search, the effort of a correct anticipation of the potential overlap of graphemic forms between different lexical items exceeds that of checking the actual results, at least in the present case, and since some rare spellings of the queried verb could also be missed, no further restrictions were imposed. For similar reasons, character combinations such as <ck> within the root were retained despite the fact that they resulted in a number of false-positive hits of *likken* ‘to lick’ (besides occasional true-positive hits in the *PCEEC*, cf. above). Most of these were easy to identify, however, based on the semantic context. Finally, false-positive hits belonging to the homophone verb *liknen* ‘to compare’ could not be avoided, since possible spellings of that verb quite plausibly present

possible spellings of *likēn* ‘to please’ as well. In such cases, syntactic complementation can be a useful criterion for the disambiguation of graphemic forms, since the argument structure of *likēn* ‘to compare’ will, by default, require the specification of a reference point as the third argument, which is typically expressed by a PP in the present context. Since allowance must be made for diachronic variation in the complementation patterns of the experiencer verb, a certain overlap between the two verbs is possible, however, so that the criterion of syntactic complementation cannot be applied too rigidly or even pre-defined in the query.

Overall, the method developed in the above paragraphs of extracting verbal forms of specific lexical items from the parsed corpora by a sample-based representation of the verbal base as combinations of sets of discrete graphemic units presents a compromise between a high detection rate and a decrease in accuracy. The former is particularly desirable in the case of limited data and justifies the additional effort required to filter out false-positive hits from the results, some of which are rather easily detected. The fundamental process of establishing the spelling variants of a lexical base derived from the *MED* quotations is not exceedingly laborious and constitutes a pragmatic approach that is individually tailored towards each item submitted for query. The ultimate success of the approach is, of course, sensitive to a given item’s susceptibility to potential homographs, which is most pronounced in cases of existing homophone verbs. Since the queried strings are based on a sample of spellings, the method is also not entirely failsafe. For example, an additional graphemic variant of *likēn* was encountered in the *PPCME2* in the form of <ly₃ken>, which was neither represented in the *MED* quotations nor anticipated in the adopted modifications of the spellings of the root. The existence of additional forms that were not retrieved from the corpus is, therefore, a serious possibility. For *likēn*, the final query yielded 256 tokens of 88 types in the *PPCEM2*, not counting the 1 additional form that was discovered later, which only occurred once. This ratio of types to tokens already implies a high degree of variation, which is likely to result, at least in part, from the presence of unrelated lexemes among the search results. Upon further inspection, 187 (73%) of the retrieved forms were identified as forms of *likēn* ‘to please’, while 66 forms (25.8%) were identified as forms of a different verb or as lexicalised participial forms, and 3 forms (1.2%) were left undecided. In the *PCEEC*, the query returned 676 tokens of 48 types, 643 (95.1%) of which were identified as forms of *likēn* ‘to please’, while the remaining 33 forms (4.9%) were identified as forms of one of three different verbs (*licēnsēn* ‘to license’, *likēn* ‘to compare’, *likēn* ‘to lick’). The accuracy of the query was thus much higher in the *PCEEC* than in the *PPCME2*, which is not surprising given the higher orthographic variability in the latter corpus. For both corpora combined, the query retrieved 932 tokens of 118 morpho-graphemic

types in total, which indicates a partial overlap of the forms in the two corpora. 830 (89.1%) of these forms were identified as forms of *liken* ‘to please’, 99 forms (10.6%) as forms of other verbs or as lexicalised participles, and 3 forms (0.3%) were left undecided. The results are summed up in table 2 below:

	PPCME2	PCEEC	combined
types	88	48	118
tokens (%)	256 (100%)	676 (100%)	932 (100%)
<i>liken</i> ‘to please’ (%)	187 (73%)	643 (95.1%)	830 (89.1%)
other (%)	66 (25.8%)	33 (4.9%)	99 (10.6%)
undecided (%)	3 (1.2%)	0	3 (0.3%)

Given that the starting point of the process of retrieval, viz. the body of quotations of *liken* in the *MED* excluding the supplementary materials, already comprises 257 instances of the verb, the return of 187 verified forms of *liken* in the *PPCME2* appears to be relatively low. If forms from the *PCEEC* in the periods M3 and M4 are included, however, the total number of ME forms returned by the query increases to 359, which already presents a significant improvement over the initial inventory of quotations. It should also be remembered that the individual parsed corpora provide a much smaller database than the corpus underlying the *MED*. If larger corpora with syntactic parsing or POS tagging were available, the results could certainly be improved even further. The great variability of orthographic conventions in the ME data accounts for the limited accuracy of the query when applied to the *PPCME2*. At least some types of false-positive hits, however, particularly those of more complex verbal stems, are easily identified by manual inspection of the results, and others, particularly those of homophone verbs, require reference to the context for detection in either case. On the other hand, it should be noted, once more, that the success of the applied method is contingent upon the properties of each specific item. In the current example case of *liken*, the query appears to have yielded acceptable results, which will serve as the database for the remainder of the study. The data retrieved for the other verbs investigated in this study was generally less copious, however. Since this fact is not specifically due to the process of retrieval, the details of their attestation will be described in the relevant parts of the results section in 4.2 below.

3.3. Subsetting

3.3.1. Non-finite Forms

The data that were extracted from the two corpora by the process outlined above contain a variety of grammatical forms of the verbs under investigation, comprising both finite and non-finite forms. For several reasons, the latter were not included in the final dataset, however. The most important reason for their exclusion is the fact that the complementation patterns of non-finite verb forms may be subject to different syntactic rules than those of finite verb forms, to which the primary focus of this study was restricted. The difference in complementation is already apparent from the fact that non-finite forms lack an explicit experiencer argument more frequently than finite verb forms, even though they are generally less frequent than the latter. Another point is that the experiencer argument is frequently realised by a PP with certain non-finite forms, especially with the present participle. Although this realisation of the experiencer argument can be compared to its less frequent occurrence with finite verb forms, and although it was generally interpreted as equivalent to oblique case forms of pronominal experiencers, the present participle, in its predicative use, is virtually indistinguishable from a participial adjective, which means that the prepositional realisation may be the result of the syntactic rules of a different lexical category rather than the indication of an impersonal construction. Also, the oblique case of experiencers in sentences with subject-to-object raising is not a reliable indicator of the construction type, since its case is governed by the matrix verb in such instances. Observations whose construction type could not be established were eventually removed from the dataset in either case, but for non-finite verb forms, this exclusion was part of a general decision. Before such non-finite instances are dismissed, a brief overview of the types of observations affected by this restriction will serve to provide an idea of the actual diversity of the data. They include observations with infinitives, present participles, gerunds and past participles, which will be surveyed and exemplified in the following sections.

3.3.1.1. Infinitive

Infinitival forms of the verbs under investigation occur in a variety of different contexts, not all of which were excluded from the investigation. One of these is the combination with (pre-)modal verbs like *mouen* ‘may’ and *connen* ‘can’. Even though these are in a phase of transition during the ME period, their conjunction with the verbs under investigation was generally interpreted as yielding complex finite forms. Contexts that were excluded from the

investigation involve instances of subject-to-object raising (or “accusive-with-infinitive constructions”) and subject-to-subject raising (or “nominative-with-infinitive constructions”), which are illustrated in (63) and (64) respectively:

(63) [...] I never saw **hym** lyke thing better , [...]

‘I never saw him like anything better’ (MORE,280.009.100)

(64) **he** Seimes to wonderfully like all our orders (FERRAR,275.021.412)

The first example shows an instance of subject-to-object raising, where, in generative terms, the experiencer argument *hym* is raised from the subject position of the subordinate clause containing the verb form *lyke* to the object position of the matrix clause, where it receives case governed by the verbal form *saw*. Although the experiencer argument is grammatically expressed, its oblique case form is not contrastive, since it does not depend on the construction of the impersonal verb. The second example presents a similar case, in which the experiencer argument *he* is the subject of the raising verb *semen*, from which it receives nominative case. In principle, the rules of case government for such instances of raising would need to be ascertained by a separate investigation that takes potential diachronic changes of the applicable syntactic rules into consideration, but since there were no contrasting examples of subjects of *semen* that received oblique case from the impersonal verb, and since all examples of this construction were restricted to the EModE period, it was assumed that occurrences of nominative experiencer subjects of *semen* were syntactically ambiguous.

Outside of raising constructions and a few other less frequent cases, infinitival forms repeatedly occur as clausal complements of adjectives and nouns. These are exemplified in (65) and (66):

(65) and , sins you weare pleased to like the other that came from thens , I venture to present this to you , and with it the best seruis of Your most louing , faithfull , and humbell seruant , Dorothe Randolph . (CORNWAL,225.139.1956)

(66) and I have just cause to like my seat wel , where I thanke my God I have an house in good forwardnes [i.e. near completion] . (HASTING,53.019.549)

In (65), *liken* functions as clausal complement of the deverbal adjective *pleased*. In (66), it complements the verbal noun *cause*. The form of the experiencer in these instances can be identified as a null constituent (PRO), which can be recovered semantically by its coreferentiality with the preceding topical element, but which lacks phonological and

morphological substance. Due to this lack of an expressed grammatical form, the relevant observations cannot be identified in terms of construction type based on the case marking of the experiencer argument, and, unlike cases with null constituents in finite clauses, there is no option of determining the construction type by reference to verb agreement, since the experiencer and the stimulus argument, whose grammatical properties in (65), for example, can be recovered as second person plural and third person singular respectively, do not control agreement of the non-finite verb form. Given the general ambiguity of observations containing infinitival forms of the verbs under investigation, these observations were consequently excluded from the dataset. Their verbal semantics remain interpretable, however, as causative ‘please’ and receptive ‘like’ (for this terminological distinction, which was introduced in section 2.1.4.3.2 above, cf. Fischer/van der Leek 1983: 341). The receptive interpretation seems to be underlying both of the above examples, since the glossing of *like* requires PDE *like* rather than *please*. Such a semantic interpretation would be consistent with a personal reading of these instances with a nominative subject. Verbal semantics alone were not taken as an indicator of the construction type, however, since they do not cover the syntactic criterion of oblique vs. nominative case of the experiencer argument. The same holds true for the other non-finite forms, which will be discussed presently.

3.3.1.2. Present Participle

The Penn-Helsinki parsed corpora generally do not attempt to distinguish between present participles and deverbal adjectival forms, since the line between the two is often difficult to draw and can require a closer examination than the general parsing process affords. This includes cases of lexicalised subsenses of the participial adjective that are more or less removed from the basic verbal meaning. For the present participle of *liken*, which is derived by two competing suffixes in Middle English as either *liking* or *likand* and their respective variants, the *MED* cites two such senses, viz. ‘in good physical condition, healthy, vigorous’ and ‘sexually aroused’, besides the primary sense ‘pleasing’ s.v. *liking(e)* (MED25565). Both of these senses actually occur in the extracted data. An example of the former is given in (67) below:

(67) Thys man , beyng in good heele of body & hys tweyn hors craske & lykand [...]

‘this man, being in good physical condition and his two horses vigorous and healthy’
(CMKEMPE,10.172)

Since the participial form *lykand* in (67) does not require an experiencer argument in this specific subsense, it was not included in the present investigation. Also occurrences of the present participle that are not as clearly distinct from the basic verb in terms of their semantics were excluded. These comprise cases of non-finite adverbial clauses and attributive and predicative participles. The first cases are similar to the infinitival clauses discussed above in that they generally presuppose a grammatically unexpressed constituent, which does not exhibit overt morphological case marking and does not control verb agreement. Participial clauses were hence not analysed in terms of their construction type. An example of such a participial clause is given in (68), adjectival forms are illustrated further below in (69) and (70):

(68) Honoured Sir , haveing received this day a letter from the navey Board with a new method of a bond for me to signe , not likeing it seemes the first , or at least thay being resolved never to leave me in quiet , but turment me , and my freinds , to the very last : I have therefore humbly made bould to give your honour more truble , (PEPYS,170.078.1229)

The text token cited in (68) contains two non-finite adverbial clauses besides the one with the participial form *likeing*, which is the main focus here, viz. *haveing received [...]* and *thay being resolved [...]*. The last of these shows that the adverbial participle can be optionally accompanied by a pronoun, which is realised by the nominative form *thay* ‘they’ in the present case, while the other two instances contain null constituents, which are coreferential with *the navey Board* and *I* respectively. While no such instances are actually attested in the dataset of *liken*, it is at least conceivable that a pronominal realisation of the experiencer argument in conjunction with a participial form could appear either in the oblique form or the nominative form, neither of which would be diagnostic of the construction type. Since no verb agreement is marked on the non-finite form either, there is thus no grammatical distinction between an impersonal construction and a transitive construction in terms of morphological case or verb agreement. The most likely construction of a finite equivalent of the present example could, of course, be inferred from other instances in the same text, if such are present, or, from instances of the same time period. Since the correspondence from which the example is taken dates to the final EModE period E3 (1640–1710), the non-finite instance is most likely paralleled by personal constructions. In order to avoid any circular inferences, however, the grammatically ambiguous instances of participial clauses were generally excluded from the dataset.

For adjectival forms, the expression of an additional argument besides the noun that the adjectival form qualifies is optional or at least not strictly required in all cases. The following examples illustrate both the attributive position and the predicative position:

(69) Ac þe leazinges likinde . byþ more grat zenne / [...]

‘but the flattering lies are a greater sin’ (CMAYENBI,63.1167)

(70) bot ay it stirres hym to do some gode þat myght be lykand til God , as in praying , or in wirkyng profitabel thynges , or in spekyng of Cristes passyon ; [...]

‘but aye it [sc. that blessed love] stirs him to do some good that might be pleasing to God, as in praying or in the working profitable things, or in speaking of Christ’s passion’ (CMROLLEP,78.249)

The postposed attributive adjective *likinde* in (69) constitutes a technical term for a particular type of sin in the religious literature, and, as such, it is more specialised in terms of its semantics compared to the basic verbal meaning. It also differs from the verb in typically leaving the experiencer argument unexpressed, since the participants that are being flattered by these lies are normally not at issue. The experiencer is also not usually expressed with less specialised instances of the attributive participle, although the specification of an experiencer argument in conjunction with the participle is an option both logically and grammatically, which is demonstrated by the second example, in which the participle occurs in predicative position following the copula *be*. This distribution of overt experiencer arguments is representative of other instances in the data, but it is not completely tied to the position of the participle. When an experiencer argument does occur, however, it consistently takes the form of a PP like *til God* in example (70) above or, more commonly, headed by the preposition *to*. The fact that nominal constituents do not occur as complements of the participial form makes it probable that the expression of the experiencer by a PP is the result of different complementation rules compared to those of finite verb forms and not simply equivalent to the oblique case of pronominal experiencers. One prerequisite of this interpretation is that instances of the participle in predicative position are, indeed, interpretable as adjectival forms and not as main verbs in grammaticalised progressive constructions with a structural auxiliary *be*, as these would count as finite forms similar to passive constructions formed with *be* in conjunction with the past participle. Apart from the consistent complementation by PPs, the lack of any indication among the examples that a grammaticalised progressive reading is, in fact, applicable presents an

argument against the latter view, however, and adjectival instances of the present participle where thus equally excluded from the dataset.

A semantic interpretation of the verbal meaning underlying present participles with null constituents or adjectival complementation patterns is possible based on their antecedent or on the constituent qualified by the participle, similar to that of infinitival forms discussed above. In (69), for example, the semantic interpretation of *likinde* must be causative, since the head noun *leazinges* qualifies as stimulus argument but not as experiencer argument in view of its semantic property of lacking animacy. The diachronic development of the semantic interpretation of participles of *liken* generally matches that of finite impersonal and personal constructions in that the recorded instances in the *PPCME2* are generally causative, i.e. equivalent to a PDE glossing by ‘please’, while later occurrences are mostly receptive, i.e. mostly equivalent to a PDE glossing by ‘like’. What is interesting is that the causative reading of the participle occurs as late as the early 17th century. Specifically, it is found in letters from collections dated to 1625 (e.g. WENTWOR,228.073.1022) and 1629 (e.g. BARRING,76.027.577), during which period finite impersonal constructions are already greatly outnumbered by personal constructions. As with the infinitival forms, however, the semantic interpretation of the present participle was not interpreted as a grammatical equivalent of different morphosyntactic construction types.

3.3.1.3. Gerund

Verbal nouns, or gerunds, are occasionally tagged as deverbal participial forms in the corpora, in which case they were detected by the script utilised to extract potential tokens of the verbs under investigation. The grammatical status of these items as nouns is established by their grammatical context, which differs from that of finite verbs in both instances cited in (71) and (72) below:

- (71) Mistris Hildersham is very desyreous , so it may stand with your ladiship 's good likeing , to have goodman Litton and his wyfe to be with her for her comfort and helpe . (BARRING,95.046.872)
- (72) Because in another consideration it hath ben thought that the Jalousie betweene England and ffrance of not liking that either should attayn anie possession in the lowe Countries maie be cause of hindrance , [...] (EDMONDE,318.017.382)

In (71), the form of *liken* is accompanied by a genitive attribute, which is a typical modifier of nouns, and in (72), *liking* itself is a complement of the preposition *of* and hence analysed as NP, although, like other verbal nouns, it preserves the ability to take a clausal complement and the negative particle *not*. Since the possessive form of the experiencer *your ladiship's* in (71) is quite untypical of the complementation patterns attested with verbal forms, instances of verbal nouns were excluded from the analysis, just like the other non-finite forms.

3.3.1.4. Past Participle

The final non-finite form to be discussed in the present context are past participles. These occur in combination with the structural auxiliaries *haven* ‘to have’ and *ben* ‘to be’ in grammaticalised perfect and passive constructions, which were analysed as complex finite verb forms. Since perfect constructions do not pose any additional problems for the analysis of the construction type in terms of personal or impersonal construction, they will not be discussed further at this point, and passive constructions will be discussed in the subsequent section 3.3.1.5. Apart from these cases, a small number of independent past participles remain, which primarily differ from passive constructions in their lack of an auxiliary verb. The experiencer argument is either expressed by a PP or absent altogether. It is also noticeable that all recorded instances of such participles of *liken* have a receptive interpretation and occur relatively late in the EModE period E2, cf. the two examples given in (73) and (74):

(73) [...] I have expected and with silent and stolne teares implored and expected reliefe and have utterly neglected or rejected all other meanes , how well liked of others soever .
(STUART,126.004.46–47)

(74) My verey good Lord , fynding some scrupulositie in men for disbursing so great a some of money now presently , as this bringer can declare , I haue offered vnto the counsell of the exchequor as much as I haue offered to your Lordship , which although very well liked of , yet they staying vntill your solliciter 's coming , I thought good to make your Lordship now privye therof ; [...] (PARKHUR,184.044.765)

In the first instance in (73), the participle *liked* functions as modifier of the preceding stimulus argument *all other means*, while the experiencer argument is realised by the PP *of others*. The fact that the experiencer is unspecific and functions as point of comparison for the means employed by the author herself provides a good motivation for this kind of construction, which, by way of its backgrounding of a participant with otherwise central involvement, is comparable

to a passive construction in terms of perspectivisation. The same effect is apparent in the second instance in (74), which equally provides additional information about the stimulus argument, viz. the offer of money under discussion, while the experiencer argument is left entirely unexpressed. Given that the realisation of the experiencer by a PP is not equivalent to the use of PPs in finite active clauses, since, in the former case, it depends on the presence of the past participle, which could not simply take a nominal constituent marked by oblique case as an additional complement, this realisation was not interpreted as a marker of the construction type in terms of personal and impersonal construction. Instances without an expressed experiencer argument were equally not analysed in this respect, and participial constructions containing an instance of the past participle were thus generally excluded from the dataset. An additional point illustrated by the second example is the elaborate style of writing, which includes many participial constructions, two further examples of which are contained in the token above, viz. *fynding some scrupulositie [...]* and *they staying [...]*. This predilection for involved hypotaxis by means of participial constructions is particularly characteristic of some of the more formal correspondence in the *PCEEC*, and it also seems to account for the individual occurrence of past participles in contexts where finite passive constructions including a structural auxiliary verb are otherwise more commonly employed. All of the non-finite forms discussed above were ultimately excluded from the dataset that served as basis for the regression model of the odds of the construction types. The following section will address passive constructions, which are similar to past-participial clauses in grammatically distancing or backgrounding one of the participants, either by realisation as a PP or by complete ellipsis.

3.3.2. Passive Constructions

The importance of passive constructions in terms of general frequency and diachronic distribution differs greatly for each of the verbs under investigation. With *liken*, for example, the distribution is equally restricted as that of the past-participle as a dependent clause, as most of the recorded instances occur during the EModE period E2. In terms of their communicative function and complementation, the two constructions are equally similar, resulting in the backgrounding of one of the participants involved in a situation. Given the late occurrence of passive constructions with *liken*, it is not surprising that the nominative subject of the corresponding active construction represents the experiencer argument in all of the recorded cases, which results in a receptive interpretation of the predicate meaning in all instances. The example in (75) will serve to illustrate this:

(75) The loftie maskers were so well lyked at court the last weeke that they were appointed to performe yt again on Monday (CHAMBER,I,498.040.1821)

More often than not, the experiencer argument is entirely absent in passive constructions like the one in (75). In other cases, it is expressed as a peripheral participant in the form of a PP headed by *of* or *by*. The function typically associated with the passive construction is clearly present in the above example. The maskers at court were generally well-liked, and there is hence no need to specify any members of the court as participants in the situation, although they are probably implicit in the locational adjunct *at court*. Since the stimulus argument [*t*]he *loftie maskers* is animate, the selection of its semantic role does not follow automatically from its semantic properties. In the present case, the immediate context is sufficient for its disambiguation, however, since the re-appointment to perform at court implies that the maskers are the stimulus and not the experiencer of pleasure, which is also part of the very nature of their profession. A second instance of the passive, in which the experiencer argument is realised by a PP, is cited in (76) below:

(76) The question is how this jorney \$will \$be lyked by my frendes .
(BACON,II,13.203.3515)

The semantic interpretation of the predicate is receptive here as well, since the conceptual starting point of the corresponding active construction would be the experiencer argument *my frendes*. The use of the passive construction inverts the position of topic and focus of the active construction, which seems to be the primary function in the above case, since the experiencer is clearly relevant to the statement. Its realisation by a PP is different from that of experiencers in active constructions, however. This is not only apparent from the fact that, despite a partial overlap in the use of the preposition *of*, which is the most frequently used preposition in active clauses, a different set of prepositions is generally used in passive clauses, most commonly the preposition *by*. A second difference is the fact that the prepositional realisation in passive constructions is not replaceable by nominal or pronominal constituents, unlike the PP counterparts in impersonal constructions. For this reason, passive constructions, like the non-finite instances discussed above, were excluded from the dataset, which was thus ultimately restricted to finite active clauses. This subset was further reduced by the exclusion of all observations that did not contain both verbal arguments, since these are required for the establishment of the dependent variable and the predictor variables. Such instances of missing experiencer and stimulus arguments will be illustrated in the following two sections.

3.3.3. Missing Arguments and Null Constituents

3.3.3.1. Experiencer

The experiencer argument of verbs of emotion can be realised by various lexical and syntactic categories, but, in some cases, the experiencer may also remain grammatically unexpressed without the option of recovery from the context. For verbs of emotion, this may seem surprising at first, since the perception of emotion logically presupposes the existence of a sentient being as experiencer of the emotion. The lack of its grammatical expression thus needs to be interpreted not as the actual lack of an experiencer but as the lack of its specification in a given context. A typical case in point are the passive constructions and certain instances of the non-finite clauses discussed above, but also finite clauses with active constructions can be missing an experiencer argument, which usually implies a general or generic kind of experiencer in such cases. An example of this construction is given in (77):

(77) & biwite mi bodi þt is al bitaht to þe from ulche fulþen . þt neauer mi sawle ne isuled beo in sunne . þurh þt licomes lust þt lutle while likeð .

and protect my body, which is all given to thee, from any filth, that never my soul not sullied be in sin by that body's lust that little while pleases

'and protect my body, which is entirely put in your charge, so that my soul never be sullied in sin by that bodily lust that pleases for a little while' (CMMARGA,57.37)

In this passage, the stimulus argument is specified by the antecedent of the abbreviated relative pronoun *þt* 'that', which refers back to *licomes lust* 'bodily lust'. The statement about the transient nature of this particular type of lust is of general validity, however, and the experiencer is thus of a general nature and remains unexpressed. The context of the present instance appears to suggest that human kind in general is implicit as potential experiencer in this statement, but a more circumscribed applicability of a general experiencer in a more specific context is equally conceivable. Such experiencers, which are neither explicitly expressed nor grammatically recoverable from the surrounding context, will be referred to as "zero experiencers" in the present context, while deleted constituents, which have a grammatical antecedent and are thus uniquely identifiable, will be referred to as "null constituents". The latter are of greater importance for the present study, since they may still contribute to the understanding of the development of personal and impersonal constructions, while the former were ultimately not utilised for regression modelling. Observations involving zero experiencers like the one above can still be interpreted in terms of their verbal semantics, however. The stimulus argument,

which is explicit in the above instance, functions as the conceptual starting point of its clause and thus requires a causative interpretation of the verb. This analysis follows from the semantic properties of the stimulus argument, since an inanimate, abstract participant like *lust* is a suitable stimulus argument but not a very suitable experiencer argument. The causative interpretation is also inherent in the glossing of *liken* as ‘please’ rather than ‘like’ in the translation above. If an experiencer argument was expressed and realised by a case-bearing constituent, it would presumably occur in an oblique case form, but since the clause effectively does not contain such a constituent, it would make little sense to assume the assignment of covert oblique case marking. The comparison with contemporaneous instances of the same time period would equally suggest a causative interpretation of the verb, since the Bodley manuscript of *St. Margaret*, from which the passage above is taken, is dated to c. 1225 and thus part of the earliest ME period M1 (1150–1250), in which the impersonal construction of *liken*, for which the causative interpretation is characteristic, is still dominant. The inference of the construction type from other observations within the same time period could lead to circularity, however, and such an approach was thus generally not employed. Since the overt or covert morphological form of the experiencer was used as the central criterion of the construction type, observations with zero experiencers were not included in the analysis.

In contrast to observations with zero experiencer arguments, observations in which the experiencer is realised by a null constituent were, in some cases, interpretable for the purposes of the analysis. The grammatical antecedent of such constituents specifies the argument semantically, which is deleted syntactically under coordination. As a prerequisite for their inclusion, null constituents potentially control verb agreement, which was used as one of the criteria of the construction type (cf. section 3.4.1 below). Their grammatical and semantic properties can be established based on the coreferential antecedent. Additional instances will be discussed further below as a distinct categorial realisation of the experiencer and the stimulus argument, but a preliminary example is already given in (78) below in order to illustrate the principal difference between experiencers realised as null constituents and zero experiencers:

(78) And , so as owyr Lord wolde , þei; obeyd not ne ___; lyked not þe menys which wer proferyd hem ,

‘And, like our Lord wanted it, they did not obey nor did they like the means which were offered to them’ (CMKEMPE,60.1343–1344)

In contrast to zero experiencers, the experiencer in (78) is recoverable from the immediate context as *þei*, whose precise identification depends on the wider context of the passage. While

the experiencer is not grammatically expressed in the same clause as the predicate, it is still specified by the context of the proposition and thus different from zero experiencers, which are entirely unspecified and rely on contextual implications for their identification. The above instance also demonstrates that isolated tokens of the verbs under investigation are not always sufficient for establishing the semantic and grammatical properties of their constituents, which is one of the drawbacks that were discussed in relation to attempts to use *MED* quotations as the sole database of a syntactic investigation. The phenomenon is particularly pronounced in cases like the one above, where a pronominal constituent with text-internal reference is deleted under coordination. With the help of the corpora, it was usually possible, however, to identify the antecedents of deleted constituents or pronominal items. These were recorded alongside the formal experiencer argument, which facilitated their inclusion in the database.

3.3.3.2. Stimulus

Like experiencer arguments, explicit stimulus arguments may be missing from an observation for a variety of different reasons. There is a fundamental difference between the prerequisites for the inclusion of observations containing missing stimulus arguments compared to that of missing experiencer arguments, however. The former merely require the interpretation of their semantic features and categorial realisation, while the latter need to be interpretable in terms of their case marking in order to facilitate the identification of the applicable construction type, whose definition was primarily based on the case marking of the experiencer. Since semantic features are interpretable also in the case of implicit constituents, a third category of missing stimulus arguments was distinguished. These will be referred to as “inferred stimuli” and essentially correspond to Allen’s NO PROP construction with an implicit propositional argument. Examples of zero stimuli, deleted stimuli, and inferred stimuli will be discussed in the following paragraphs.

The first example in (79) introduces a potential instance of a zero stimulus, i.e. the complete lack of a stimulus argument, both in grammatical and semantic terms:

(79) Now þe name of Ihesu es noghte elles bot þis gastely hele . Whare-fore it es sothe þat þay say , þat þar may na man be safe bot if he lufe & lyke in þe name of Ihesu ;

‘Now, the name of Jesus is nothing else but this spiritual health, wherefore it is true what they say that no man may be safe unless he loves und likes in the name of Jesus’
(CMROLLTR,45.932)

The stimulus argument of *lyke* in this example appears to be not only grammatically absent from the clause, but also not specifiable by reference to an antecedent or to the general context. Such cases of zero stimuli are not very common for a verb like *liken*, since the verbal semantics typically express the emotional stance of an experiencer towards a specific stimulus rather than an independent emotional state, which is found in English adjectival predicates like *to be happy* or *to be afraid*. A situation like the one above, where the manner of liking, viz. liking in the name of Jesus, takes precedence over the expression of a specific stimulus, presumably does not obtain very often. Incidentally, the experiencer argument is equally missing in this example, but it is recoverable from the preceding clause *bot if he lufe* as a deleted anaphora of the coreferential subject pronoun *he*. It is precisely this coordination with the predicate *loven* ‘to love’ that motivates the interpretation of the deleted constituent as the experiencer argument of the semantically related predicate *liken* in its receptive sense, although a contrastive use of *liken* in its causative sense ‘to please’ would be conceivable as well. Under the latter interpretation, the deleted constituent would be analysed as the stimulus argument, and the example would present a case of a zero experiencer instead, which could be identified in terms of a general experiencer like the one in the example of bodily lust in (77) above. Given the relative uncertainty of its interpretation, the exclusion of this example from the dataset seems advisable either way.

There is another context in which zero stimuli occur with some regularity, however, and that is their parenthetical use in some of the more formal letters in the *PCEEC*. An example of this usage is given in (80) below:

- (80) Sir , if it lyke your Grace , at my retorne whan I spake with the King , his Grace was very ioyfull , that notwithstanding your so continuall labors in his maters in which he saied ye haue many moo than appere to theym that see you but at Westminster or with the counsaile [that] your Grace is so well in helth , [...] (MORE,157.002.11)

The conditional clause that contains the predicate *lyke* is not syntactically integrated into the surrounding sentence, and it is not semantically integrated with the context either, since the reported information concerning the king’s sentiment about the addressee’s health does not require their approval. Instead, it serves as a kind of preface to that statement, asking leave of the addressee to deliver said piece of information, which appears to be offered *pro forma* as part of the general conventions of courtesy at the time. The alternative to assuming that the stimulus argument is missing would be to supply a stimulus argument like *that I say the following* or *to hear the following* to complete the parenthetical use of *liken*, which appears to

function primarily as a discourse-organising element. Other observations of *liken*, particularly the formulaic expression of a type to be discussed in more detail in section 3.4.1.2.2 below, explicitly contain a predicate that connects the impersonal verb with the message that is being introduced, and these could serve as a parallel on which the inference of a stimulus argument like the ones suggested above could be based. Since it is not contained in the actual observation, however, and since it cannot be inferred syntactically from the limited context of the parenthetical statement, this approach was not followed with *liken*, and the relevant instances were analysed as cases of zero stimuli instead. Even so, zero stimuli remain a fairly uncommon phenomenon for *liken*. Since such instances do not exhibit the properties that function as predictor variables in the present study, observations with zero experiencers were excluded from the analysis.

The second category of missing stimulus argument involves instances of deletion, in which the properties of the stimulus can be recovered from its antecedent. The stimulus argument is thus not actually missing in such cases, and its inclusion in the analysis is less problematic, since the presence of overt case marking as a criterion for the construction type is not as crucial as for experiencer arguments. The deletion of such constituents takes place in coordinated clauses, but also missing relativisers were analysed as null constituents. An example of the latter type is given in (81):

(81) there are some Story's in't ___ **you** will like I beleeve . (OSBORNE,21.009.383)

The semantic properties of the null relativiser in this example are identified by its antecedent *some Story's*, and the instance is thus interpretable in terms of the predictor variables, although the dependent variable construction type remains ambiguous in this particular instance due to the morphological ambiguity of both arguments and the lack of distinctive verb agreement. Other cases of deletion usually involve a coreferential subject constituent, but there appears to be at least one instance of deletion in the data of *liken* in which a deleted constituent functions as object. It is cited in (82) below:

(82) and for y=e= cloth y=t= Mr Simpson sent me a while ago; **I like** ___; very well ,

‘and regarding the cloth, that Mr Simpson sent me a while ago, I like [it] very well’
(FLEMING,303.121.2024)

If the lack of an anaphoric pronoun in the relevant clause is not merely an unintended oversight by the author, the deleted stimulus argument, which is recoverable from the topic-introducing PP as y=e= cloth y=t= Mr Simpson sent me a while ago, functions as object of *liken*, since the

experiencer argument *I* is marked as subject by nominative case and control of verb agreement. Regardless of the unusual type of deletion, the instance is thus interpretable in terms of its construction type, and since the properties of the stimulus are identifiable based on the antecedent of the deleted constituent, this instance, like other instances of deletion, were included in the dataset.

The final type of missing stimulus argument involves instances that will be referred to as inferred stimuli. As the name suggests, these are not grammatically expressed but inferred from the context of a given verbal token. They are similar, in principle, to instances of deletion, since the relevant participant is semantically specified, except that the missing argument is equivalent to a full proposition rather than an individual participant and does not always have a direct grammatical antecedent. They differ from zero stimuli in that a specific stimulus is always implied rather than a general stimulus or the absence of a stimulus altogether. Such instances were thus classified a separate case. Consider the example in (83):

(83) And when he herd Crist preche , **hym** lyket soo well þat he fatte Seynt Petyr , his broþer , to Crist forto here his prechyng .

‘And when he heard Christ preach, he liked it so well that he brought Saint Peter, his brother, to Christ to hear his preaching’ (CMMIRK,6.130)

This example can be interpreted in several different ways. One interpretation would be to assume deletion of the anaphoric pronoun *it* in the matrix clause *hym lyket soo well*. Such a constituent would not have a direct antecedent in the preceding adverbial clause, however, and thus cannot be said to be coreferential with a preceding element. Alternatively, the provision of pronominal *it* could be viewed as a semantically empty expletive subject, which is apparently not required by the syntactic rules that apply to the above example. If the subject constituent were semantically empty, however, and the proposition equivalent to ‘he was (generally) pleased’, one could also argue that the absence of the subject reflects the lack of a specific stimulus argument, i.e. a zero stimulus. This is clearly not the case in terms of semantics, however, since the embedded adverbial clause indirectly states the stimulus argument, viz. hearing the preaching of Christ. The grammatical form that this stimulus argument would most likely take in the matrix clause seems to be that of an infinitival clause *to hear Christ preach*, which captures all participants of the proposition including the PRO subject of the infinitive, which is coreferential with the experiencer argument *hym*. Coreferentiality with the topical constituent is, in fact, the main criterion by which the non-finite form of the inferred stimulus

was identified as opposed to a finite subordinate clause, which would be equally possible in the above example.

A second example, in which the infinitival form is explicitly stated in the context, is given in (84) below:

(84) Ne I seye nat , ther as I speke of divisioun of confessioun , that if thou have licence for to shryve thee to a discreet and an honest preest , where **thee liketh** , and by licence of thy curaat , that thow ne mayst wel shryve thee to him of al alle thy synnes .

‘nor do I say, where I speak of the division of confession, that, if you have licence to make confession to a discreet an honest priest, where it pleases you, and by licence of your curate, that you may not make confession to him of all your sins’

(CMCTPARS,324.C2.1567)

Again, the stimulus argument is not grammatically expressed in the clause that contains the verbal form *liketh*, but it can be inferred from the superordinate clause as *to shryve thee to a discreet and an honest preest*. The subject of this infinitival stimulus is coreferential with the topical experiencer argument *thee*. An instance in which the subject of the inferred stimulus is not coreferential with the experiencer argument is given in (85):

(85) " [...] & , 3yf it lyke 3ow , I wyl gon a-geyn to hir for recorde þat I mevyd no sweche mater . "

“and if you like, I will go to her again for the record that I moved no such matter”

(CMKEMPE,133.3130–3132)

In this case, the stimulus argument of *lyke* can be inferred in the form of a finite subordinate clause [*that*] *I wyl gon a-geyn to hir for recorde [...]*, whose subject is not coreferential with the experiencer argument *3ow* ‘you’. The instance also differs from the previous one in that it contains the expletive subject pronoun *it*, which anticipates the clause from which the stimulus argument is inferred. This information is not a direct grammatical equivalent of the stimulus argument, however, since it constitutes the matrix of the sentence, while the stimulus requires the form of a subordinate complement clause. The main point, in any case, is that instances of inferred stimuli, just like instances of deleted stimuli, were included in the final dataset, since they are part of the semantic predication and interpretable in terms of their semantic properties, while zero stimuli, for which this is not the case, were ultimately removed from the dataset. In terms of their syntactic category, inferred stimuli were identified with the respective categorial realisation in which they were inferred, viz. infinitival and finite subordinate clauses. This

presents a certain inconsistency, since deleted arguments were otherwise analysed as a separate category of null constituents, but given the differences between such deleted arguments and the inferred stimuli discussed above, this approach seemed a viable compromise. The issue will be taken up again in the relevant section on the categorial realisation of stimulus arguments.

3.3.4. Ambiguous Observations

A final point that needs to be addressed in the context of subsetting of the dataset are observations whose expression of the dependent variable could not be disambiguated by morphological case marking or any of the other criteria used to establish the construction type of impersonal verbs. The application of these criteria and the type of observations that remained ambiguous will be discussed more fully in the relevant section on the dependent variable construction type (cf. section 3.4.1 below), but it should already be noted at this point that ambiguous observations, despite the fact that they could not be utilised in the analysis of potential semantic or syntactic influence on the dependent variable, are sometimes interesting in their own right, since the question of structural ambiguity is one of the most contended issue regarding the presumed actuation of reanalysis. While ambiguous observations were thus excluded from the dataset on the basis of which the regression model was built, they were not simply discarded but evaluated separately for the information they provide, especially in the case of verbs whose data were insufficient for logistic regression.

In some cases, ambiguity of an observation also arises from the heterogeneous realisation of a predictor variable. This was not encountered very frequently in the data, but some instances of coordinated stimulus arguments did not coincide in their semantic properties and thus could not be identified in terms of the respective variable. For example, if one of two coordinated constituents was human and the other constituent non-human, this observation was not interpreted in terms of animacy. Alternatively, two separate syntactic structures, one for each stimulus argument, could be assumed, and these could then be counted as two separate observations of a given verb. This was not done for coordinated constituents with identical properties, however, which means that a twofold interpretation of coordinated constituents with different properties would have been inconsistent. A simple interpretation according to the properties of either one of two coordinated constituents would have conflicted with the theoretical premise that different semantic properties have different effects on the construction type. Due to the low number of cases affected, their exclusion seemed unproblematic in any case. For the same reason, an additional level of mixed animacy was not assumed either, since

this level would not be very well represented in terms of frequency, and its expected effect on the dependent variable would not be very informative.

3.4. Annotation of Variables

The principal layout of the present investigation includes four predictor variables besides the dependent variable construction type, viz. animacy, abstractness, category and periods. Of these, animacy and abstractness are semantic variables, which derive from the semantic properties of a given constituent's referent, while category reflects the syntactic realisation of that constituent. The periods refer to the date of the manuscript or letter from which the data are taken. In the following section, the motivation of the distinction of different variable levels will be discussed for each of the predictor variables. This is necessary because semantic variables like animacy and abstractness can be conceptualised and subdivided into discrete categories in a variety of ways. The identification of observations with these variable levels will be illustrated for selected instances, as these do not always fall clearly within either one of the categories. The principal aim is to provide transparency about the decisions that were made during the coding process, which is susceptible to subjectivity, but also to provide further examples of the data as a complement to their quantified treatment in the remainder of the study. The realisation of different variables is therefore illustrated for both the experiencer and the stimulus, even though the properties of the latter argument are most relevant to the analysis. The main exception is the expression of the dependent variable construction type, which is defined primarily in terms of the morphological case of the experiencer argument. Since morphological case is not overtly marked on all types of experiencer arguments, the construction type is based on a broader construct, which includes references to covert case and constituent order. The criteria for identifying the construction type will be the topic of the following section.

3.4.1. Construction Type

As discussed in section 2.1.1 of the introduction, impersonal constructions were defined more narrowly in the present context as constructions that involve an experiencer argument marked by oblique case. This definition allows for the general inclusion of *ME liken*, which frequently occurs in type-ii constructions with a second complement in the nominative case, and it avoids the more controversial definition of impersonal constructions as subjectless. The central

criterion for determining the construction type of the verbs under investigation is thus case marking. According to the semantic hypothesis which the present study proposes to test, the development of the experiencer's case marking is influenced by its semantic interpretation as more or less agentive. This influence can only be indirect, of course, since the notion of agentivity has no immediate expression by any distinct grammatical category in the grammar of English. Instead, the increasingly relevant grammatical relations expressed by constituent order and case marking constitute syntactic configurations that are assigned to semantic arguments based on a hierarchical selection process. Nevertheless, the prototypical mapping of certain semantic roles onto certain grammatical relations results in characteristic patterns, which can arguably influence the assignment of case to experiencer arguments based on their relative identification with more prototypical agents. The following subsection 3.4.1.1 discusses the observability of overt case marking of the experiencer, while the subsequent section 3.4.1.2 discusses the establishment of covert case marking of this argument based on several additional criteria.

3.4.1.1. Case Marking

3.4.1.1.1. Pronouns, Nouns, Adjectives

Case marking continues to be relevant in the context of experiencer arguments during the ME and EModE periods, since it never entirely disappears from the domain of personal and anaphoric pronouns. The continued distinction of these items between nominative and oblique case forms can serve as an immediate indicator of the relevant construction type according to the definition of impersonal constructions as constructions that involve an oblique experiencer argument. A morphosyntactically related phenomenon is the realisation of core participants by PPs, which becomes more prominent during the ME period. Even though such constituents do not receive case marking as a whole, PP experiencers were generally regarded as structural equivalents of the oblique pronominal forms which they replace, at least in the context of finite active clauses. The situation is different for NPs filled by full nouns or a morphosyntactically range of uninflected pronouns like *which* or *that*. The former generally lose their overt case marking by the late ME period in all morphological cases except the genitive singular. During earlier stages of Middle English, the dative singular ending *-e* could still contrast with an endingless nominative-accusative form in former *a*-stems such as *ston* 'stone' (cf. nom.-acc. sg. *ston* vs. dat. sg. *ston-e*), but since nouns derived from more complex stem formations in **-o-*, **-ja-* and **-an-* already ended in *e* in their basic form (cf. e.g. *soule* 'soul', *ende* 'end',

name ‘name’), their equivalent dative singular endings in *-e* and *-e(n)* were usually not distinctive. The loss of *e* in final syllables and some analogical restructuring ultimately led to a merger of the different paradigms into a single declension with endingless forms. In writing, however, final *e* was, conversely, extended to the basic form of nouns like *stone*, which makes it even less transparent as an indicator of case marking (cf. Mossé 1952: 47–49; also Fisiak 1968: 80f.). A rare example of a case-marked nominal form encountered in the *PPCME2* is *drihtne* from *The Lambeth Homilies* cited in (86) below:

(86) ah dele we ure ehtan mid wisdome . swa þet hit **drihtne** likie .

but deal-out(SUBJ) we our wealth with wisdom so that it the-Lord(DAT) please(SUBJ)

‘but let us deal out our wealth wisely so that it may be pleasing to the Lord’
(CMLAMBX1,105.972)

In this case, syncope of the stem vowel *e* from the basic form *drihten* ‘Lord’ suggest that the final *-e* is an actual ending and not merely a graphemic variant. Apart from such rare instances, however, case-marked nouns were generally not encountered in the corpora, which means that nominal experiencer arguments did not provide morphological indicators of their construction type. Adjectives and demonstrative pronouns (including the definite article) did not compensate for this lack, since they only exhibit vestiges of case marking from the earliest ME documents onwards (cf. Mossé 1952: 47–64f. and 60f., Fisiak 1968: 83 and 90). The disappearance of case marking from nominal declensions and their modifiers is particularly relevant for the analysis of the correspondence data from the *PCEEC*, since certain types of correspondence exhibit a preponderance of nominal forms of address such as *your grace* and *your mastership* in favour of pronominal forms, and these do not provide morphological clues regarding the construction type. A general exclusion of instances with morphologically ambiguous experiencer NPs from the study was judged undesirable, however, since the amount of available data was already limited by the size of the dedicated resources and the requirement of subsetting the data to finite active clauses. Additional criteria for determining the construction type of the extracted instances of impersonal verbs were therefore applied whenever possible. Before turning to these criteria, some additional remarks on the case marking of pronouns are required, since, even in this domain, not all morphological distinctions are preserved during the period under investigation.

3.4.1.1.2. *Ye/You* Contrast

As is well known, the English second-person singular pronoun *thou* with its contrasting oblique form *thee* was restricted to informal contexts and otherwise replaced by the second-person plural pronoun *ye* with its contrasting oblique form *you* (earlier *eu*, *eo*, cf. Mossé 1952: 54f.). The latter contrast is largely lost during the early EModE period, however, with occasional instances of the oblique form in subject function registered as early as the 14th century (cf. Nevalainen 1996: 65). Since this development directly affects the crucial distinction between nominative and oblique experiencers, and since the use of the second-person plural pronoun as a polite form is particularly prominent in the correspondence of equal members of the gentry, which constitutes a large part of documents contained in the *PCEEC*, all instances from the two corpora involving forms of the second-person plural pronoun were checked for the distinctiveness of their case forms. In order to account for potential intra-individual variation, all applicable author-recipient combinations in the *PCEEC* were checked individually, since it seemed conceivable that a son or daughter, for example, might address their parents by a different choice of forms than they would address their siblings. The process thus primarily involved an examination of pronoun usage within a given letter or within letters of the same author-recipient combination if the first specimen did not yield sufficient diagnostic contexts. In some cases, this criterion had to be relaxed, however, and judgement was based on letters by the same author to other recipients or on letters by other authors within the same collection of letters. One reason for this was that some authors, despite a general effort of the compilers to include at least ten letters per writer (cf. Nevalainen/Raumolin-Brunberg 1996: 42), are simply not represented extensively enough to yield an informative sample with the above restriction in place. This is particularly true for cases where the consistent use of a form in either subject or object function needs to be established, since this can only be done by way of generalisation, while inconsistency in usage can, in principle, be demonstrated by a single counter-example. If a given sample contained only the oblique form of the pronoun and this form only occurred in object function (or as complement of a preposition), this did not yet constitute probative evidence of its distinctiveness if contrastive contexts in which the nominative form would be required were absent. Another point is that, even in cases of extensive representation of an author, the available evidence was partly limited by the fact that polite forms of address such as *your lordship* and *your grace* frequently stand in place of pronominal forms in certain parts of the correspondence. As a result, the classification of some of the instances of pronominal forms as distinctive or non-distinctive was necessarily tentative, although the majority of cases were decided with confidence.

While no pronounced differences between specific author-recipient profiles were established during the investigation, a certain degree of variability was encountered not only within given collections of letters from a single time period, but also in individual authors. Thomas More, for example, uses both *you* and *ye* with consistency in a letter to Thomas Wolsey dated 1519 (MORE_002), but in another letter to Thomas Cromwell dated 1534 (MORE_030), there is at least one example of the apparent extension of *you* to subject function, cf. (87):

(87) but I besech yow for our Lordis love , that yow be not so wery of my most cumberouse suit , but that it may lyke yow at such oportune tyme or tymys as your wisdom may fynde , to helpe that his Highnes may by your goodnes be fully enformed of my true faithfull mynde , and that in the mater of that wykked woman there neuer was on my parte eny other mynd than good , (MORE,500.030.458)

In this example, the oblique form *yow* is used as subject in conjunction with *wery* ‘weary’ as subject complement. While the difference between the two letters is probably not yet indicative of a difference in addressing two different recipients, or at least not of one that could be established within the confines of the present study, it illustrates the existence of variation in More’s pronoun usage and highlights, once more, the difficulty of establishing valid generalisations based on a limited sample of forms. The investigation of multiple samples from the same author in the form of different author-recipient combinations thus also counteracts the general problem to some extent, since the convergence of several samples will yield a more solid basis of evidence. Divergent samples, on the other hand, were assessed independently in cases like the one above. The occurrence of *you* was deemed distinctive in the first letter and non-distinctive in the later letter, since the first letter contained multiple instances of both *you* and *ye* in conservative usage, and the probability that only the occurrence with an impersonal verb was innovative seemed rather low in the given context. The example also illustrates that the mere occurrence of the nominative form *ye* is not enough evidence to prove contrastiveness of the oblique form *you*, since also the second letter contained a couple of instances of *ye* in subject function besides non-distinctive *you*. Similar facts were encountered in some of the *Cely Letters* (1474–1488) and the *Letters of William, Lord Paget of Beaudesert* (1547–1563), where *ye* continued to be restricted to subject function, while *you* occurred as either object or subject. Also the nominative form was occasionally encountered in object function, cf. the following example from the *Correspondence of Robert Dudley* (1585–1586) in (88):

(88) And thus having bin long , as I \$can \$not {TEXT:cannott} others chuse having so much to imparte to you , I wyll byd ye farewell , longing styll to hear from ye .
(LEYCEST,76.020.807)

In view of such evidence, the distinctiveness of each morphological form had to be assessed independently, and no conclusions about the morphological contrast could be drawn from the simultaneous occurrence of the two morphological forms.

The general picture of the contrast between *ye* and *you* was more clear-cut than in the above examples, however, as most letters either displayed distinctive use of the two forms or the complete extension of *you* to all contexts. Letters with predominantly distinctive use include *The Letters and Accounts of William Brereton of Malpas* (c. 1520–1539), the *Life and Letters of Thomas Cromwell* (1523–1540), the *Original Letters, Illustrative of English History vol. 1* (c. 1418–1529) as well as most instances from the larger collections of the *Paston Letters and Papers of the Fifteenth Century* (c. 1425–1519) as well as *The Stonor Letters and Papers* and the *Supplementary Stonor Letters and Papers* (c. 1420–1483). The dating of these collections illustrates that personal correspondence could exhibit the conservative use of the two forms as late as the 16th century, which seems to qualify the assumption that correspondence is generally more likely to adopt linguistic innovations compared to other written genres. Interestingly, only a single exception to conservative usage was encountered among the many different author-recipient combinations sampled from the last-mentioned collection, which involves a letter written to Elizabeth Stonor by Goddard Oxbyrge, a family servant, cf. (89):

(89) and , Maistes , pleasthe you to undirstond that I have \$Received a letter ffrome you by Davy Wrixham , the which letter I undirstod wele , and schoche matter ase you he [i.e. have] wreten to me (STONOR,II,8.045.728)

This instance can perhaps be taken as an example of socially conditioned variation, since no comparable instances were encountered in letters written by members of the Stonor family. At the same time, other servants represented in the collection displayed conservative usage of the morphological forms in their letters as well, so the above example should probably not be overemphasised. The instances of *ye/you* from the PPCME2 were generally found to occur in contexts of distinctive usage, which include *The Brut or the Chronicles of England* (c. 1400), Chaucer's *The Tale of Melibee* (c. 1390), Malory's *Morte Darthur* (a1470) and Caxton's *History of Reynard the Fox* (1481). At the other end of the spectrum, instances of the generalisation of *you* are dominant in *The Papers of Nathaniel Bacon of Stiffkey* (c. 1569–1594), the *Barrington Family Letters* (1628–1632), *The Oxinden Letters* (1607–1642) and *The Letters of Lady Arbella*

Stuart (1588–1611) besides many others. These are generally dated later than the collections with distinctive usage and continue well into the late 17th century, which shows that the diachronic development of the replacement of *ye* is well underway during this period. An approximate cut-off point that could be used for a broad classification might be the middle of the 16th century, but given the previously discussed variation within the collections and within certain authors, the individual examination of occurrences of *ye/you* during this transition period offers a more precise assessment. A similar investigation was conducted for the relative pronoun *who/whom*, whose forms likewise require the confirmation of their distinctiveness in a similar period of transition. The extent of their occurrence as experiencer arguments was much more limited than that of the personal pronoun, however. While a more thorough investigation dedicated specifically to such morphological contrasts might yield different results on individual points, the procedure was deemed necessary in the present context in order to establish the case marking of pronominal experiencer arguments in all available cases. Additional criteria for establishing the construction type of impersonal verbs will be discussed in the following sections.

3.4.1.2. Other Criteria

Since observations with morphologically marked experiencer arguments constitute only a subset of the extracted data, it was deemed necessary to rely on additional criteria in order to secure as broad a database as possible for the quantitative analysis of the diachronic development of the construction types. In the case of *liken*, for example, more than one third of the experiencer arguments are realised by caseless constituents, and these could not enter into the analysis if case marking was used as the sole criterion for the construction type. Despite the fact that impersonal constructions were defined as constructions involving an experiencer argument marked by oblique case, overt case marking thus was not the only criterion applied in determining the construction type of impersonal verbs. One such additional criterion was the order of constituents in clauses containing expletive *it* (or earlier *hit*). This approach is easily justifiable, since caseless constituents demonstrably exhibit particular manifestations of case in certain environments when replaced by case-bearing constituents. Even though a given NP constituent may not be marked by case, one can argue for the existence of underlying covert case marking based on instances in parallel constructions, which exhibit overt case marking in the case of most pronominal constituents. The question is, of course, which manifestation of case marking can be postulated for a given environment. Since word order is still relatively free

in Old English and only gradually takes on a predominantly grammatical function during later stages, an all-encompassing account of word order types and their relation to the diachronic development of impersonal verbs would introduce an additional unknown into the analysis rather than clarify the role of covert case marking. Expletive *it*, on the other hand, arguably has the effect of disambiguating the grammatical role of a given constituent by occupying the appropriate syntactic slot. For this reason, the constituent order in so-called *it*-constructions can be used as an indicator of the applicable construction type. As will become apparent below, the evidence of *it*-constructions can also be transferred to certain cases of formulaic expressions, which do not contain an instance of the expletive themselves, but which allow for the syntactic identification of the constituents based on structural similarities. Such formulae were equally used as additional criterion for establishing the construction type of impersonal verbs. The third criterion takes the form of indirect evidence from the case marking of the stimulus argument. In cases where the experiencer bears no overt case but the stimulus does, it can be reasonably assumed that the case of the two constituents will not be identical. Overt nominative case marking of the stimulus argument thus can be interpreted as an indication of covert oblique case of the experiencer argument and vice versa (possible objections to this criterion based on so-called double-object constructions will be discussed below in section 3.4.1.2.3). The final criterion employed in the present context is evidence from verb agreement. Based on the assumption that verb agreement is controlled by nominative constituents, constructions whose constituents exhibit contrastive grammatical properties allow for the identification of the controller and, by extension, of the nominative constituent even in cases where neither constituent is overtly case-marked (possible objections based on instances of non-standard verb agreement will be discussed below in section 3.4.1.2.4). All of the above criteria were applied in sequential order and only to the subset of data not previously disambiguated. For example, the criterion of constituent order in *it*-constructions was only applied to observations that did not already contain a case-marked experiencer argument, the criterion of formulae was only applied to observations that did not already constitute an interpretable instance of *it*-constructions and so forth. The order was chosen in order to facilitate the annotation process, applying informative criteria that were easily observed before less informative criteria that were more difficult to establish. Since a significant diachronic increase in the use of expletive *it* can be observed with some impersonal verbs at least, this criterion ranked highest wherever present, and its application to the establishment of the construction type of impersonal verbs will therefore be the topic discussed first in the following sections.

3.4.1.2.1. Expletive *It*

The increasing use of expletive constituents in the history of English affects impersonal verbs to varying degrees. While not all of the verbs under discussion take part in the development, it is very noticeable for a verb like *liken*, where expletive *it* and its equivalent forms (earlier *hit*, occasionally also *that*) occur in a total of 241 or about 29% of all 831 observations. Their distribution will be discussed in more detail below, but the most important point in the present context is that many instances of the verb which lack overt case marking of the experiencer can be disambiguated in terms of their construction type by reference to the relative word order of the constituents in constructions with an expletive, since these can be compared directly to equivalent constructions containing case-marked experiencer arguments. In order to facilitate this process, the surface constituent order of all observations was analysed and generalised into different types of *it*-constructions. Adjuncts, for example, were not represented in the constituent order types, since they only affect the absolute position and not the relative order of the constituents in relation to the verb. Auxiliaries and lexical verbs were recorded separately, but the classification into constituent order types was based on the lexical verb alone. Since auxiliaries and lexical verbs commonly occur adjacent to each other and do not affect the relative word order of the other constituents, this rarely poses a problem except in the relatively infrequent cases of hortative clauses with a clause-initial auxiliary, in which the expletive intervenes between the auxiliary and the lexical verb (*may it like you ...*). Such instances were recorded schematically as *it_V_E* based on the position of the lexical verb, even though comparable auxiliary-less hortative clauses (*like it you ...*) display initial position of the lexical verb. With 210 instances, *it_V_E* was also the most frequently encountered word order type among the 241 observations of *liken* containing an expletive. An overview of these types and their frequency is given in table 3 below:

type	<i>it_V_E</i>	<i>V_it_E</i>	<i>E_V_it</i>	<i>it_E_V</i>	<i>E_it_V</i>	<i>it_V</i>	total
frequency	210	21	6	2	1	1	241

The stimulus argument, which is realised as a finite or non-finite clausal complement in the majority of cases involving an expletive, typically follows the experiencer, but in relative clauses and subordinate interrogative clauses, clause-internal constituents of the stimulus may be fronted together with the *wh*-pronoun or the relativiser as in (90):

- (90) Elys your man may tary sumtyme after your lordships departing with whom it may lyke **you** to leave your commission (PAGET,11.001.6)

Even though part of the stimulus argument *with whom ... to leave your commission* is fronted, its position in relation to the lexical verb of the matrix and to the experiencer argument is equivalent to that of other clause-final stimuli in *it*-constructions, and since this position is constant across all examples, it was not recorded explicitly in the constituent order types. Another possibility is that the stimulus is not expressed at all and either inferable from the context or simply left unspecified as a general stimulus. These cases are likewise subsumed under the constituent order type *it_V_E*.

If the case marking of the experiencer argument in the most frequent word order type *it_V_E* is considered more closely, it becomes apparent that 102 (48.6%) of all 210 instances are overtly marked by oblique case, while overt nominative case marking does not occur on any of the experiencer arguments. Another 12 instances (5.7%) are realised by PPs, and 1 instance (0.5%) was classified as uncertain, since the distinctiveness of *you* could not be decided for that case. This strongly suggests that the remaining 95 instances of caseless experiencer arguments (45.2%) can be interpreted as constituents with covert oblique case, and that the constructions in which they occur can be considered impersonal constructions. The fact that overtly marked constituents display oblique case is, of course, not very surprising, since the word order type *it_V_E* is a manifestation of the increasingly dominant SVO word order. The expletive fills the subject position and the experiencer fills the object position, in which it receives oblique case. The postverbal position of the experiencer in these cases contrasts with impersonal constructions without an expletive, since these often display preverbal position of the experiencer (cf. section 2.1.1.1.4.1). In both constructions, case-marked experiencer arguments are marked by oblique case, however, despite the fact that nominative case marking presents a syntactically viable alternative. In terms of this criterion, constructions with and without an expletive are thus both interpretable as impersonal constructions. Similar facts about case marking can be gathered for the second most frequent word order type *V_it_E*, which differs from the first in that it displays inversion of the expletive and the lexical verb. Out of 21 instances, 13 (61.9%) contain an oblique experiencer argument and none contain a nominative experiencer argument. It is thus reasonable to attribute covert oblique case to the 8 remaining caseless constituents (28.1%). Not all instances with expletive *it* can be automatically interpreted as impersonal constructions, however. An example is the instance with *it_E_V* order in (91) below, in which both the expletive and the experiencer argument precede the verb:

(91) | & all þiss iss bilumpenn me , || Swa summ itt **Drihhtin** likeþþ , |

and all this has befallen me so as it the-Lord pleases/likes

‘and all this has befallen me as it pleases the Lord / as the Lord likes it’
(CMORM,I,166.1371)

This instance is arguably compatible with both an impersonal and a personal reading, which is reflected in the glossing of the verb as either ‘please’ or ‘like’. Pronominal *itt* could either be in the subject position of an SOV construction, but it could also be a preposed pronominal object. There are two further instances of this constituent order type in the *PPCME2*, and both exhibit overt oblique case marking of the experiencer. This is not a sufficient basis for the classification of the above example, however. There is another constituent order type in which both the experiencer and the expletive occur preverbally, albeit in reversed order, viz. E_it_V. Here, the experiencer is likewise marked by oblique case, which demonstrates that OSV is a possible word order in *it*-constructions. Within a conservative implementation of the criterion, the above example thus remains ambiguous. Finally, the word order type E_V_it presents two different cases, which are separated by a temporal gap and a difference in construction type. The *PPCME2* contains 2 instances, both of which exhibit an experiencer marked by oblique case, while all four instances from the *PCEEC* contain an experiencer marked by nominative case. An example of each case is cited in (92) and (93) below:

(92) Or ellus an-oþer manere if þe like it betere

Or else another manner if thee like(SUBJ) it better

‘Or else another manner, if you prefer’ (CMHORSES,115.297)

(93) I like it not amis , that you haue stayed his younger brother , if he will doe
(PASTONK,81.051.973)

In both instances, the verb is constructed with an additional adverb, which promotes the extraposition of the clausal stimulus argument, which is explicitly expressed in (93), but inferred in (92). The contrast in case marking of the experiencer argument presents a good example of the diachronic change under investigation. At the same time, the example in (93) demonstrates that expletive *it* can fill the object position of a clause, and that the presence of an expletive and the constituent order need to be considered in combination in order to classify observations of the verb as instances of impersonal or personal constructions. If both factors are taken into account, many of the instances that lack overt case marking of the experiencer

can thus be disambiguated in terms of their construction type by way of the syntactic configuration which the expletive establishes.

The inclusion of observations with caseless experiencer arguments ensures a broader database for the investigation of the diachronic change from impersonal to personal constructions, and since case-bearing experiencer arguments continue to be marked by oblique case in most instances of *it*-constructions, the extension of their analysis to caseless constituents in demonstrably comparable cases is plausible also from a definitional point of view. It should be borne in mind, however, that constructions with expletive *it* constitute a specific subtype, which typically occurs with clausal complements or with other heavy constituents such as NPs involving an attributive relative clause. The stimulus argument in these cases will thus be propositional and invariant with regard to animacy in most cases, while cases with different and more variable kinds of stimuli cannot be disambiguated by the same criterion and thus remain potentially excluded from the analysis. The subsumption of both subtypes under a single construct of impersonal constructions will be addressed below, and the implications of the inevitable skewing of the utilisable data will be discussed for each verb that this applies to. On the other hand, limiting the discussion to instances without expletive constituents would affect not only the cases with morphologically ambiguous experiencers that were discussed above, but also those in which the experiencer is case-marked, and since these constitute a significant share of the total data for certain verbs, this option seemed less desirable. A second general point to note is that the impact of *it*-constructions on the total data partly depends on the fact that they occur in formulaic expressions with a relatively high token frequency. These exist in variant forms either with or without the expletive, and the fact that the expletive is partly optional in such formulaic expressions is another argument in favour of the combined treatment of such constructions. It also suggests that the analysis of *it*-constructions can be extended further to expressions without an expletive if these are not already disambiguated by case-marked experiencers. The application of this criterion will be the topic of the following section.

3.4.1.2.2. Formulaic Expressions

While a comprehensive definition of linguistic formulae is beyond the scope of the present study, the concept can be intuitively characterised as involving the co-occurrence of certain invariable components with a significantly above-chance frequency. Allowing for variable degrees of schematicity of these components facilitates a conceptualisation of language in which practically every grammatical structure can be conceived of as a formula or construction.

This is a central point in variously formalised versions of Construction Grammar (cf. e.g. Goldberg 1995, Barðdal/Gildea 2015). For the present purposes, a more pragmatic definition of formulaic expressions was applied, however, which will serve as an additional diagnostic tool in the classification of the construction types. The applicability of this criterion is generally much more limited than that of the *it*-constructions discussed above, since all of the formulaic expressions discussed at least optionally include an expletive and were thus frequently already disambiguated by the former criterion. At the same time, the main reason for considering them as an additional criterion was their structural similarity with *it*-constructions.

Two formulaic expressions were recognised in connection with *liken*. Only one of these included instances without an expletive, to which the analysis of instances with an expletive was extended. The other formula always contained expletive *it* and is discussed for expository purposes only. The schematic structure of the first expression is given in (94):

(94) like (it) E (to) wit / know (that) ...

The expression consists of the finite verb in initial position followed by an optional expletive, the experiencer argument, and an infinitival stimulus complement involving a knowledge verb and an embedded clausal complement, which is optionally introduced by the complementiser *that*. The function of the expression is to introduce new information to the discourse, which is sometimes made explicit by the use of sentence adverbs like *furthermore*. The verb occurs in the third-person singular subjunctive, which has a hortative function, but there are also instances of what appear to be third-person singular indicative forms as in (95):

(95) And fferþer more , liketh it **your good ladyschipe** to wete that this same daye I come to London . (STONOR,II,53.067.1189)

The constituent order in this expression corresponds to the constituent order type V_it_E, for which covert oblique case marking of the experiencer was established above. This analysis was extended to cases without an expletive like the one given in (96):

(96) \$After my most speciall recomendacion , like **your maisterchip** wete that the office is taken at Ocle jn lyke fourme as Suthwell can schew you , (PASTON,II,220.368.9734)

Since the experiencer in the above example is realised as a NP, the construction type cannot be established based on overt case marking, and since the criterion of constituent order was generally used for instances with an expletive only, the example cannot be disambiguated by that criterion either. Instead, it is analysed as impersonal based on its structural similarity with the previously disambiguated instances of the formula. The quantitative significance of

formulaic expressions without an expletive was, in fact, rather minor. In the case of the present formula of *liken*, only seven observations were analysed based on this criterion, while all other occurrences had been previously disambiguated by either case marking or constituent order in *it*-constructions. Most of these occurrences also conformed rather closely to the schema in (94), but, in some cases, the expression was used with other verbs of cognition like *to remember* or with nominalised variants such as *to have knowledge*.

The degree of variability of the stimulus argument was more pronounced in the second formulaic expression, which is similar to the first in including *to wit* and *to know* as possible predicates, but which generally involved a broader range of verbs. Due to this variation, the expression is represented with a higher degree of schematicity in (97):

(97) it may like E + inf.

In contrast to the first formula, the second expression conveys a hortative sense by means of the auxiliary *may* rather than a third-person singular subjunctive, and expletive *it* is present in all instances. Since the word order type *it_V_E*, to which the expression belongs, was previously analysed as an impersonal construction, all instances whose construction type was not already disambiguated by case marking were classified based on this criterion. In functional terms, the formula appears to primarily fulfil a discourse function, since the predicate of the infinitival stimulus includes frequent occurrences of *to understand* and *to be advertised*, which signal that new information will be conveyed by the complement clause of the stimulus. An example of the expression is given in (98):

(98) Hit may ferther lyke your good Grace to vnderstand that the Kingis Grace mych alloweth your prudent answeare made vn to th'Emperors Embassiator vppon the saufconduicte . (MORE,276.007.72)

Other instances of the expression include various nominal references to cognition such as *to call to your gracious remembrance*, but also a range of other verbs such as *to declare*, *to examine* or *to send*. The latter expand the assumed formula to such a degree that they could also be simply thought of as instances of productive language usage, and whether they are ultimately interpreted as instantiations of higher-level schemata depends on one's theoretical orientation. The first components of the expression, however, which also include the frequent realisation of the experiencer as *your (good) Grace* or similar titles, are rigid enough to justify the recognition of a formulaic expression for the present purposes.

As a final point, it should be noted that the proportion of formulaic expressions in the data becomes rather substantial if all instances of the two aforementioned formulae are counted and not only those in which they were the decisive criterion for the classification of the construction type. In addition to the 7 instances of *liken* classified as impersonal constructions by application of the criterion, an additional 39 observations were recognised as instances of formula 1, and an additional 59 observations as instances of formula 2, which yields a total of 105 or 12.6% of all 831 observations of *liken*. The proportion is even higher when directly compared to the subset of 761 finite active clauses, where formulaic expressions constitute 13.8% of all observations. While the inclusion of formulaic expressions may have a negative impact on the discernibility of synchronically active factors in the choice between two construction types, it does increase the database of analysable constructions and appears to be representative of the type of language used in the sources. Their exclusion, on the other hand, would be near-coextensive with the exclusion of *it*-constructions, since most instances of the formulae contain expletive *it*. Two additional criteria for the construction type, the case-identity constraint and verb agreement, will be discussed in the subsequent sections.

3.4.1.2.3. Case-Identity Constraint

There are various references in the literature to the fact that impersonal verbs occasionally occur with two oblique constituents in Middle English (cf. e.g. Jespersen 1984: 221, van der Gaaf 1904: 68). An example of this occurrence was already given in (20), repeated here as (99):

(99) how likes **þe** me nowþe

how pleases(3-SG) you(OBJ-SG) me(OBJ-SG) now

‘how do you like me now’ (*Will. of Pal.*, 1740)

Both the experiencer argument *þe* and the stimulus *me* in this so-called double-object construction are marked by oblique case, and neither of them controls verb agreement of the third-person singular form *likes*. In this respect, double-object constructions are superficially similar to “true” impersonal verbs like OE (*of*)*hreowan*, which occurs in type-i constructions with an oblique experiencer and a genitive stimulus argument. Since *liken* in the above example is a typical type-ii verb, however, its use with two oblique constituents is clearly not an archaism. On the whole, double-object constructions appear to be rather sporadic occurrences during the ME period, and they were never encountered in the present data except for one instance of the verb *reuen* given in (100):

(100) **me** reweþ Marie þi sone and þe .

me causes-pity, Mary, thy son and thee

‘I feel pity, Mary, for thy son and thee’ (CMEDVERN,257.723)

In this example from a version of the *Speculum Edmundi* (c. 1390), both the experiencer argument *me* and one of the coordinated elements of stimulus argument, *þe*, exhibit oblique case. Since *reuen* ultimately continues the type-i verb *hreowan*, this usage might well be considered conservative. It could also be derived from the Latin exemplar, however, of which the text is a relatively close rendering according to the bibliographical information provided in the corpus. In either case, the construction is not very typical of the verb, which is otherwise constructed with a nominative NP stimulus or a PP stimulus as in (101):

(101) | **Himm** reoweþþ ec off alle þa | | þatt follþhenn deofless lare , | | [...]

Him rues also of all those that follow the-devil’s teaching

‘He also pities all those that follow the devil’s teaching’ (CMORM,I,193.1583)

Although the above examples in (99) and (100) illustrate that certain exceptions do exist, the basic assumption that the experiencer argument and the stimulus argument receive non-identical case marking remains valid as a general constraint. As a result, the case marking of the stimulus argument, whenever it is morphologically unambiguous, can supply indirect evidence of the case marking of the experiencer argument, which can be analysed as nominative in the case of oblique stimuli and as oblique in the case of nominative stimuli. PP stimuli as in (101), on the other hand, cannot be used as diagnostic tools, since they are found to occur with either nominative or oblique experiencers.

The role of the case-identity criterion in expanding the total database is rather moderate, since it was only applied to those instances that had not been previously disambiguated by case marking of the experiencer, constituent order in *it*-constructions or the identification with a formulaic expression. In the case of *liken*, 20 observations were classified by means of this criterion, and all of these, except 2 early instances with nominative stimuli, were classified as personal constructions based on overt oblique case marking of the stimulus argument. Given that all of the personal instances belonged to the EModE period (E1 = 1 instance, E2 = 4 instances, E3 = 13 instances), their analysis as personal transitive constructions rather than potential double-object constructions can be established with great certainty. Unlike the constructions with expletive *it*, which were predominantly constructed with inanimate

propositional stimuli, these instances expand the evidence of animate stimuli in the database, albeit mainly for the later periods, during which personal constructions are virtually generalised. The distribution of observations disambiguated by the final criterion of verb agreement was not as restricted, and the application of this criterion will be discussed presently.

3.4.1.2.4. Verb Agreement

Assuming that verb agreement is exclusively controlled by nominative constituents in Middle English and Early Modern English, its evidence can be used to identify the covert case marking of morphologically ambiguous experiencer and stimulus arguments. If the experiencer controls verb agreement, it can be identified as covertly marked by nominative case in a personal construction, and if the stimulus controls verb agreement, the experiencer can be identified as covertly marked by oblique case in an impersonal construction based on the case-identity constraint described above. Contrary to initial expectations, verb agreement was not a particularly powerful criterion for establishing the construction type of impersonal verbs, and it was therefore employed last in cases in which all other diagnostic criteria had proved inconclusive. One reason for its limited usefulness in disambiguating the construction type is the fact that verb agreement itself is not overtly marked in several grammatical categories. This generally applies to past-tense forms of the verb, but also to present-tense forms in the subjunctive mood, which is much more widespread in Middle English and Early Modern English, since it is generally required in conditional clauses and also occurs in several types of main clauses in order to express different types of modality. Another point is that many of the emerging modal auxiliary verbs such as *can* and *may* lack certain markers of verb agreement in the present tense on account of their derivation from earlier preterit-present verbs. Their use increases significantly during the EModE period and results in a high frequency of cases to which the criterion of verb agreement cannot be applied. Even existing markers of agreement are not always distinctive, since an ending like *-(e)th* may be used for both the third-person singular indicative and the plural in some dialects. Such endings therefore require an analysis of the surrounding context in order to establish their distinctiveness, similar to the pronominal forms *ye/you*. Finally, the criterion obviously requires non-identity of the agreement features of the experiencer and stimulus argument. Two third-person singular constituents, for example, cannot be distinguished by verb agreement, even if this is distinctively marked on the verb. Whether combinations of experiencers and stimuli with different agreement features are likely to occur partly depends on the semantics of a given verb, of course. For *liken*, the stimulus

argument was predominantly in the third-person singular (717 or 86.3% of all 831 observations), which means that third-person singular experiencers, which were likewise the most frequent type (320 or 38.5% of all observations), had a relatively low chance of being disambiguated by verb agreement. This also holds true for observations with PP stimuli. Even though these do not control verb agreement themselves, an impersonal interpretation would require a third-person singular form of the verb by default, which only contrasts with the agreement features of the experiencer if the experiencer is not also third-person singular.

After the exclusion of previously disambiguated observations and those in which the experiencer and stimulus argument exhibited identical or not distinctively marked agreement features, only 18 instances of *liken* remained that were classifiable based on verb agreement. 2 instances displayed agreement with the stimulus, which was interpreted as an indication of an impersonal construction, and 16 displayed agreement with the experiencer, which was interpreted as an indication of a personal construction. As with non-identical case marking, there are occasional exceptions to the expected form of verb agreement, potential examples of which were already given in (18–20) above. An additional example is presented in (102):

(102) Most deare Sister , I have receved that money which you was pleased to send me by
M=r= Greenhill , which came soe conveniently as I could wish .
(CORNWAL,265.169.2322)

In this instance, the third-person singular form *was* fails to agree with the grammatical features of the second-person plural pronoun *you*, and although the clausal complement *which ... to send me by M=r= Greenhill* could, in theory, control third-person singular agreement, this would be most unusual in the present case. Instead, the instance can either be analysed as a passive construction of *plesen* or as a copular construction involving a deverbal adjectival form of *plesen* as a predicative complement (such cases of verbal and adjectival participles are generally not distinguished according to the annotation practice of the Penn-Helsinki corpora). Since the criterion of verb agreement was applied infrequently overall, the chances of basing a classification on spurious instances nevertheless remain rather small. At the same time, verb agreement has the advantage of being applicable to deleted constituents, which could not be disambiguated by case marking. One such instances is given in (103):

(103) Sonne , I have receyved your letter (BACON,I,69.051.922) and do like well your
diligence . (BACON,I,69.051.923)

Since the auxiliary *do* would be marked for third-person singular if it agreed with the stimulus argument *your diligence*, control of verb agreement must be exercised by the experiencer *I* in this example, which is deleted under coordination in the relevant clause. In this way, it was possible to classify the observation in terms of its construction type even though neither the experiencer nor the stimulus argument is overtly marked by case. The following sections will outline the annotation of the semantic and syntactic properties of the experiencer and the stimulus argument, which provide the basis of the investigation of their hypothesised influence on the development from impersonal to personal constructions. Specifically, these involve the animacy, abstractness and category of the arguments, each of which will be discussed separately for the experiencer and the stimulus argument.

3.4.2. Independent Variables

3.4.2.1. Animacy

As discussed above in section 2.2.2.1.4, animacy is considered to be one of the main observable factors in assessing a constituent's potential to be interpreted as agentive, and its coding is therefore of particular importance for the present study. For the experiencer argument of verbs of emotion, animacy is, of course, a logical prerequisite, since the sensation of emotion is not conceivable without the presence of a sentient entity. Little variation is thus to be expected for this argument in terms of animacy, and the variation that does occur with these verbs is certainly not enough to impact the dependent variable construction type in any demonstrable way. Nevertheless, some remarks about the different levels of animacy encountered with experiencer arguments will be of interest, since these include several types that are distinct from prototypically human experiencers. Also the stimulus argument, which is not limited by the same restrictions and theoretically able to display the full range of animate and inanimate participants, will be exemplified in the current section. This will not only facilitate the discussion of animacy as a predictor, but also provide a more thorough acquaintance with the data that were extracted from the two corpora.

The two levels of animacy that were distinguished in the present study are human vs. non-human participants. This distinction is ultimately data-driven and was considered to be the most useful one during the coding process, since non-human animate stimuli usually patterned with inanimate arguments in their semantic interpretation. Animals, for example, are conceptualised as commodities or food more frequently than as active participants, hence their

degree of animacy was deemed to be comparable to inanimate rather than to human participants. Alternatively, a distinction between animate vs. inanimate participants would have been conceivable, or even a tripartite distinction between animate-human, animate-other and inanimate participants. The latter distinction would have been too fine-grained, however, for the limited amount of data to yield interpretable results for each factor level. Since the classification of specific examples according to different categories may be susceptible to a certain degree of subjectivity, the exposition of representative examples in this section is also intended to establish maximum transparency on how the data were coded. The different levels of animacy of the experiencer argument, in particular, are primarily included for descriptive purposes, while the predictor variables intended for analysis in section 4 typically involve properties of the stimulus argument only.

3.4.2.1.1. Experiencer

The majority of experiencer arguments in the data are straightforwardly identified as human referents, which is partly a result of the fact that humans are the most common discourse topic in both the correspondence corpus and the *PPCME2*. There are certain cases, however, which do not readily correspond to the prototypical case of human experiencers. These can be roughly subdivided into four different types, viz. body parts, animals, abstract concepts and anthropomorphic beings. While body parts are clearly animate, they are not capable of experiencing emotions outside of the context of the organism to which they belong, and their occurrence as experiencer arguments thus usually signifies a part-to-whole relationship. A somewhat exceptional example of a body part functioning as independent experiencer occurs in a passage from the *Trinity Homilies* cited in (104) below, where *wombe* ‘stomach’, which is the antecedent of the experiencer argument *hire* ‘her’, is said to derive pleasure from the labour of the other organs:

(104) and alle hem þincheð to litel gief hit **hire** wel licað .

and all them thinks very little if it her well pleases

‘and they [sc. the limbs] consider everything [sc. their labour] very little if it pleases it [sc. the stomach] well’ (CMTRINIT,181.2493)

The image in this passage could be interpreted as an instance of personification, since the stomach and the other organs are described as autonomous entities, but at the same time, they clearly remain in the physiological domain, viz. the internal microcosm of the body. An

identification with human experiencers thus seems unwarranted, while an analysis as inanimate would clearly be counterfactual. As mentioned above, the distinction of a third level of animacy to represent non-human animate referents would be possible, in principle, but no such distinction was maintained for the relevant stimulus arguments, since these were generally not interpretable in terms of a higher degree of agency than inanimate stimuli, and the distinction was not introduced for experiencer arguments either, as these displayed little variation to begin with. Body parts such as *wombe* were therefore simply coded as non-human referents, and the same label was applied to other instances of body parts such as *þe heorte* ‘the heart’ and *þe flesch* ‘the flesh’, which are used metonymically to represent parts of the experiencer that are conceived of as the general seat of emotion or as particularly susceptible to certain kinds of emotion. An example of this more typical use is presented in (105):

(105) amidde þe redunge hwenne **þe heorte** likeð . kimeð up andeuociun þt is wurð
Monie bonen .

amid the reading when the heart likes / pleases, comes up a-devotion that is worth many
prayers

‘during the reading, when it pleases the heart, a devotion arises that is worth many
prayers’ (CMANCRIW-1,II.211.3021)

In this passage, *þe heorte* represents the part of the body in which the emotion is experienced. One could argue, therefore, that it actually signifies a location, which is promoted to the role of experiencer in absence of an expressed human experiencer. Full constructions with a separate locative constituent are, in fact, documented with impersonal verbs like OE *ofhreowan* ‘to cause pity’ (cf. e.g. [*m*]ec æt heortan gehrew ‘I repented at heart’, cited and discussed by Pocheptsov (1997: 474f.)), and they are also conceivable for a verb like *liken*. In the present context, however, body parts that occurred in place of a human experiencer were interpreted as non-human experiencer arguments in their own right.

The second type of non-human experiencer arguments is represented by animals, which occasionally occur as experiencers in the *PPCME2* in allegorical tales or in fables such as that of *Reynard the Fox*. In such cases, they generally constitute instances of personification, since they are presented as acting and behaving like humans in virtually all respects, cf. e.g. (106):

(106) the foxe saide / lief neuewe / how lyke ye my chyldren rosel and reynerdyn

“the fox said: ‘dear nephew, how do you like my children Rosel and Reynard’”
(CMREYNAR,57.504)

The experiencer argument *ye* refers to the main protagonist's nephew, who is asked to comment on the state of his cousins' upbringing. This statement would be indistinguishable from human dialogue if it was not introduced by *the foxe saide* or if it was not for the general context of the tale. The level of animacy can thus be considered equivalent to that of human experiencers, even though the fox and his family members are not strictly speaking human. Outside contexts of personification, animals occurred much more rarely, and these were categorised as non-human experiencers.

In certain contexts, abstract nouns occur as experiencer arguments as well. Two main cases can be distinguished, viz. abstract nouns which occur as personified concepts similar to the personified animals above, and abstract nouns which are used with reference to human experiencers as formal terms of address. Instances of the former case are found in the literary texts of the *PPCME2*, where concepts such as *warschipe* 'vigilance' and *scadwisnesse* 'reason' occur as human-like characters, cf. e.g. (107) below:

(107) Swiðe wel qð warschipe likeð **us** þt tu seist .

very well quoth vigilance pleases us that thou sayest

'What thou sayest pleases us very much, quoth Vigilance' (CMSAWLES,182.231)

Since Vigilance, the gatekeeper, acts as a character on a par with human experiencers, it was categorised as human in terms of its animacy. Instances of formal address, on the other hand, are predominantly found in the *PCEEC*, where abstract terms such as *your Lordship* and *your Grace* frequently occur in reference to the addressee. Although such abstract terms, in their literal meaning, represent inanimate concepts with a low degree of animacy compared to human experiencers, their use as terms of address, which are directed at human referents and which often are highly conventionalised, presents an argument in favour of their categorisation as human experiencers. Some instances could also be interpreted as metonymy similar to the body parts discussed above, especially if they are not conventionalised, cf. e.g. (108) below:

(108) [...] Wheryn I beseche your maistership that if my frendes pay you accordyng to my wrytyng , that ye than vouchsaf to do the said plegges be sent hider to me by such conduyte as **your wisdom** like to advise [...]

'wherein [sc. in which letter] I beseech your mastership that, if my friends pay you according to my writing, that ye then vouchsafe to cause the said pledges to be sent hither to me by such means as your wisdom may please to devise' (PASTON,II,91.285.8186)

While the first term *your maistership* in the above extract is a frequent term of address used to refer to human referents, the second term *your wisdom*, although it could be used similarly, is more likely to refer to wisdom as a specific aspect of the addressee, which can, of course, only function as the experiencer of emotion by extension. Both cases are different from human experiencers expressed by concrete nouns or pronouns in terms of the abstractness of their lexical means, but for abstract terms of address, this difference does not affect the animacy of the referent, which is clearly human. The metonymic cases, on the other hand, were analysed as containing non-human experiencers, analogous to the body parts in the earlier section, since they similarly constitute parts of the organism to which the experience of emotion ultimately needs to be ascribed, albeit in non-physical but spiritual terms.

Finally, a number of instances of experiencer arguments were classified as human experiencers based on their anthropomorphic conceptualisation rather than their actual identification as human beings. Two common examples, given the religious nature of much of the Early-English literature, include *God* and the *Holy Spirit*. An example of *ure drihten* ‘our Lord’ functioning as experiencer argument is given in (109):

(109) [...] he [sc. Job] was admōd on worde . and rihtwis on dede . and godfruht on þonke . and loðles mid alle . and þefore likede ure drihten swo wel þat he herede him ouer alle men ; [...]

‘he was humble in words and righteous in deeds and God-fearing in his thoughts and innocent in all, and therefore [he] pleased our Lord so well that he praised him above all men’ (CMTRINIT,167.2268)

The classification of God and the Holy Spirit as human rather than non-human experiencers seems unproblematic, since God created man in his image according to Christian belief, and God’s manifestation as the Holy Spirit, despite lacking corporeal form, is most readily analysed in terms of the same kind of anthropomorphism. Furthermore, literary characters such as the *horrible geauntes* ‘horrible giants’, which feature in early parts of *The Chronicles of England* and which belong to the realm of mythology, were categorised as human experiencers based primarily on their human form.

In summary, it can be stated that the number of cases in which the animacy of experiencer arguments was interpreted as human despite certain differences from prototypical instances of human beings was rather low compared to the number of unambiguously classifiable human experiencers. On the other hand, experiencer arguments that refer to physical

body parts or to character traits of the experiencer rather than the experiencer in its entirety as well as animals outside of their personified occurrence in fables were categorised as non-human experiencers. Despite the subsumption of both non-human animate experiencers in the case of body parts and animals, as well as, strictly speaking, inanimate experiencers in the case of abstract notions such as wisdom in its literal sense, these remain a clear minority compared to human experiencers. The two different levels of animacy of the experiencer cannot, therefore, have any demonstrable influence on the construction type of impersonal verbs. The above discussion of selected examples nevertheless highlights some of the difficulties in coding the semantic property of animacy in particular cases, in which intercoder agreement might not always be achieved. Since differences in coding directly impact the results, transparency about the invisible decisions that were made appears to be of importance for the procedure. This holds equally true for the coding of the properties of the stimulus argument, whose different levels of animacy will be discussed in the next section.

3.4.2.1.2. Stimulus

The Stimulus argument of the verbs under investigation is generally less restricted in its semantic features than the experiencer argument, since its endowment with perceptive qualities does not constitute a logical requirement of the predicate. With regard to animacy, this means that the full range of animate and inanimate constituents can theoretically be expected to occur with verbs of emotion and other experiencer verbs. The difference in potential to be realised by inanimate participants also implies that utterances containing a single inanimate participant are readily disambiguated in terms of their semantic roles, independent of case marking and word order, since inanimate participants will most likely perform the role of stimulus. Utterances with two animate participants, on the other hand, require the disambiguation of their semantic roles by the given context if morphological and syntactic cues are absent. An example of an inanimate stimulus argument is given in (110):

(110) [...] " I see wel , dame Prudence , that by youre faire wordes , and by youre resouns that ye han shewed me , that the werre liketh **yow** no thyng ;

“I do see, Dame Prudence, by your fair words and by the reasons that you have shown me, that you do not like the war at all” (CMCTMELI,235.C1.699)

The animate participant is most plausibly interpreted as the experiencer argument and the inanimate participant *werre* ‘war’ as the stimulus argument in this passage from Chaucer’s *Tale*

of *Melibee*, regardless of their relative position to the predicate and the distinctive oblique case marking of the experiencer argument *yow*. Note also that the impersonal reading hinges entirely on the distribution of the semantic features and the context, since a personal transitive reading would be equally compatible with the grammatical features. Other inanimate stimuli include concrete things like the ones in (111), but also finite and non-finite clauses, which will be discussed further below.

(111) Honored Cozin , I have rd [i.e. received] the gloves and ring you sent and doe like of them both exceedingly well . (OXINDE,I,292.165.2961–2)

Both the more abstract concept of war in (110) and the commodities referred to in (111) constitute stimuli that are inanimate and non-human, but instances of stimulus arguments that are non-human and animate also occur. These are typically represented by animals, cf. the specimens in (112) below:

(112) she hath wodkocs and fesants and turkes bakt [i.e. baked] and does like them so but no other way . (CONWAY,291.072.2021–2)

The woodcocks, pheasants and turkeys in this example also illustrate the context in which such non-human stimuli normally occur in the data, viz. as food or, in the case of horses and cattle, as assets that are primarily viewed as possessions rather than volitional agents. In light of this fact, it seemed more appropriate to group these stimuli together with inanimate referents and contrast both subgroups with human stimuli, for which the implication of a higher degree of agentivity can be assumed more readily, even though it is, of course, also possible to objectify human beings just like the animals in the above example. It should also be pointed out that the extent to which the property of humanness increases the likelihood of an agentive interpretation of the relevant participant is partly verb-specific, since certain verbs of emotion, for example *reuen* ‘to regret, feel pity’, take human participants as their second argument in some of their senses without any increase in agency being necessarily implied (cf. section 2.2.2.1.4.3). An example of a human stimulus argument of *liken*, for which an agentive interpretation is semantically plausible, is given in (113):

(113) My most honored Lord , I am right gladd to heare by Mr Greene , of your Lordships good health, as also to understand by your letter the good acceptance and approbacon yow haue of his seruice ; neither doe I doubt , but the more yow shall know him , the more **yow** will like him . (WENTWOR,183.052.674–5)

The construction in this example is, in fact, personal, since the experiencer appears in the non-distinctive pronominal form *yow*, for which covert nominative case must be assumed according to the case-identity constraint, while the stimulus appears in the oblique form *him*, yielding a regular transitive construction. Again, this interpretation rests entirely on the context, since both participants are animate and thus qualify for either the role of experiencer or the role of stimulus. The stimulus argument itself is clearly human in this example, but, as with experiencer arguments, there are occasional cases of non-human participants that were identified as human stimuli, either because they represented instances of personification or because they constituted abstract nouns used as formal terms of address. An example of the former are the fox cubs in the fable of *Reynard the Fox*, which was already cited in (106) above.

3.4.2.2. Abstractness

The second semantic property that was coded as indicator of a potentially agentive interpretation is abstractness. This property is, to some extent, correlated with the property of animacy, since human referents will typically also be concrete. Exceptions to this correlation may occur in the form of collective nouns such as *parliament* or *ministry*, since these can be used to imply their individual members, which are clearly human, while the collective as a whole, especially when viewed in terms of its function, is arguably less concrete than the individual members. Non-human referents, on the other hand, can be both concrete or abstract. Within this distinction, concrete referents are assumed to be more likely candidates for an agentive interpretation than abstract ones, although the collective nouns given above are, again, exceptions to this rule. It should also be pointed out that the impact of the property of concreteness on a given participant's potential to be interpreted as agentive will differ depending on the realisation of the second property of animacy. Concrete stimuli like the ones in (114) are a case in point:

(114) **the gentlewoman** neither liked gowne , nor peticoate so well , as somm little bunch of Rubies to hang in hir eare , or somm such dafte toy . (STUART,194.029.699)

The gown and petticoat in this example are concrete but at the same time inanimate, which means that they are certainly not prototypical agents, but their concreteness nevertheless makes them conceivable as a causative force that impacts the experiencer in a more tangible manner than abstract notions. A third level of abstractness was recognised for stimuli that take the form of a non-finite or finite subordinate clause, which semantically corresponds to a proposition.

On the scale of abstractness assumed here, such propositions are taken to be even more abstract than abstract nouns, since they refer to entire situations, which may involve several internal participants, none of which acts directly upon the experiencer. Three levels of abstractness were thus annotated in the data, viz. concrete, abstract and propositional participants. The typical realisation of propositional participants by clausal constituents and the fact that the syntactic realisation of the arguments was coded as a separate variable result in a considerable overlap between this level of abstractness and the associated syntactic category. The high frequency of propositional stimuli with certain verbs also implies a great preponderance of this variable level with the affected verbs. The instances that lie between the two extreme cases of concrete human stimuli and abstract propositional stimuli are not always easy to judge, however, and some examples of both experiencer and stimulus arguments will hence be discussed in the two following sections.

3.4.2.2.1. Experiencer

The general limitations of experiencer arguments in terms of animacy also hold for their association with different degrees of abstractness. Human or animate experiencers, which constitute the vast majority in the data, are normally concrete, and the predicate meaning of experiencer verbs naturally favours a concrete entity as undergoer of the experience. The few deviant cases of abstract experiencers that can be observed are those in which abstract notions were interpreted as features of the experiencer functioning as experiencer arguments in their own right. Such abstract experiencers include expressions like *youre noblesse* ‘your nobility’, *youre goodnesse* ‘your goodness’ and *your wisdom* ‘your wisdom’, the last of which was already cited in (108) as an example of a non-human experiencer. On the other hand, abstract nouns functioning as conventionalised terms of address, including *his highness* and *your grace*, were interpreted as concrete experiencers based on the semantic properties of their referents. An example of this case is given in (115):

- (115) Hit may lyke **your good Grace** to be aduertised that I haue received your Gracis lettre to me directed , wrytten the ii=de= day of Septembre , and with the same the lettres congratulatory by your Grace devised in the Kingis name to the Duke of Venice . [...]
(MORE,282.010.117)

The analysis of such constituents as concrete is consistent with their interpretation as human experiencers in terms of animacy, and since semantic features are of primary interest for this

investigation, the properties of the referent are a more suitable basis for its classification than the lexical categories used to establish this reference. At the same time, the fact that concrete referents are often referred to by abstract, reverential terms in parts of the data is of interest in itself, and this information is lost when such arguments are analysed as concrete. A separate binary factor that encodes the presence of abstract reference was therefore added to capture the information. This feature was narrowly defined as the combination of a possessive determiner with an abstract noun referring to a human participant, a pattern most commonly used as a term of address for second-person interlocutors but also applicable in cases of third-person reference as shown by the example *his highness* already mentioned above. Since abstract reference is only present in certain parts of the data, this feature cannot be used to explain the variation of the dependent variable construction type throughout the dataset. It can, however, be used to gather additional information about the type of observations in which it is present.

Concrete experiencers that are not simply human or ambiguous in some other way include body parts like *þe heorte* and *þe flesch* and a few other instances like *þe saule* ‘the soul’ and *hir felawshep* ‘her company’. The former were interpreted as non-human experiencers, since they are, strictly speaking, only parts of the experiencing organism, but in terms of abstractness, they are clearly concrete elements endowed with a corporeal form. This feature is lacking in the case of *þe saule*, but even though the spiritual essence of a human being could be thought of as an abstraction of its transcendental features, the sum of these features is probably best analysed as a concrete entity. The same is true of *hir felawshep* in (116), whose interpretation partly depends on the context in which it appears:

(116) Aftyward , whan it lykyd **hir felawshep** , þei preyde þis worthy doctowr to dyner .

‘Afterwards, when it pleased her company, they asked this worthy doctor to dinner.’
(CMKEMPE,63.1414)

In the given context, *hir felawshep* functions as a collective term for a group of people, and even though the assumption of a collective already constitutes a kind of abstraction, this instance was analysed as concrete. The same term in a different context could also be used to refer to the abstract notion of another person’s presence, in which case it would be abstract, although this would render it a rather unlikely candidate as experiencer argument. Overall, experiencer arguments were most frequently identifiable as concrete participants, and only a few items were identified as abstract. Given this low degree of variation in experiencer arguments, their abstractness was not considered as a predictor of the construction type. Of

more immediate relevance for the regression analysis are the different degrees of abstractness exhibited by the stimulus argument, which will be described in the next section.

3.4.2.2.2. Stimulus

The stimulus argument of impersonal verbs is comparatively free not only in terms of the different degrees of animacy but also in terms of the different degrees of abstractness it is able to exhibit. Given this greater potential for variability, the criteria for a given stimulus argument's analysis as either concrete, abstract or propositional require further elaboration. The first point to note is that multiple factors arguably influence the concept of abstractness. The ones primarily considered in the present context are individuation, specificity and endowment with a bodily form. Individuated entities are more easily conceptualised as concrete than collective ones, since collectives constitute an abstraction, even if their individual members are concrete. Specific items are likewise more easily conceptualised as concrete than generic ones, since generic terms such as the names of species or different occupations but also other nouns used for a class of referents constitute conceptual generalisations, while specific items refer more immediately to concrete instantiations of that concept. Finally, while endowment with a bodily form doubtlessly favours the conceptualisation of an item as concrete, spiritual entities like the soul or the Holy Ghost, which figure in the religious literature of the *PPCME2*, were equally classified as concrete.

A large number of instances were straightforwardly classifiable as either concrete, abstract or propositional. Instances of the first type include specific individual objects such as *ðat wastme* 'that fruit', which occurs in the story of Adam and Eve (CMVICES1,119.1464), but also classes of objects as, for example, *such looking glasses as you sent me*, which are requested in a business letter in the *PCEEC* (CHAMBER,1,574.046.2066). Abstract stimuli occur repeatedly in the *PPCME2* in the context of moral judgements as in (117) and (118):

(117) and halt gode þeawes þe **gode** likiað **and monnan** .

'and maintain good manners, which please God and men' (CMLAMBX1,111.1058)

(118) Ðe man slapeð faste . þe lið on swilche sinnes þe **him** puncheð swete . and wel likeð .

'the man sleeps tight who lies in such sins that appear sweet and please him well'
(CMTRINIT,201.2796)

In both of the above examples, the stimulus argument refers to a general kind of moral or immoral behaviour, and the *sinnes* ‘sins’ referred to in (118) are particularly frequent as abstract stimuli in the *PPCME2*. The third level of abstractness is represented by propositional arguments. These include constituents realised by finite and non-finite subordinate clauses as well as pronominal items with text-internal reference to such constituents. An example of the latter is given in (119):

(119) *Das þe me þincþ þu wilt godes lore bliðeliche understonden and liernin ; and ðat me likeð swiðe wel [...]*

that which me thinks thou wilt God’s lore gladly understand and learn; and that me pleases very well

‘It seems to me that you gladly want to understand and learn God’s lore, and that pleases me very much’ (CMVICES1,47.532–533)

The formal stimulus argument *ðat* in this passage refers back to the complement clause *þu wilt godes lore bliðeliche understonden and liernin*, which constitutes a full proposition that involves several participants. Such propositional arguments, which occur in a variety of different forms, were generally considered abstract, since they denote complex situations, and with some of the impersonal verbs under investigation, they constitute a particularly large share of the total number of abstract stimulus arguments.

One type of stimulus that is less easily classified in terms of abstractness is represented by indefinite pronouns and free relative clauses without an antecedent. These are maximally unspecific and could imply both concrete and abstract referents. An example from the *Chronicles of England* is given in (120), where Earl Thomas of Lancaster addresses Sir Andrew in conclusion of his reproach of the treason committed by the latter:

(120) And now goþ and do what **3ow** gode likes ;

‘And now go(-PL) and do whatever pleases you well’ (CMBRUT3,218.3910–1)

The implication appears to be that Sir Thomas, who has shown himself to be untrue to his own word, cannot be trusted, and that the speaker therefore surrenders all hope of influencing Sir Thomas’s actions. Although he might have something specific in mind when he tells Sir Thomas to do what he likes, this is not stated in the utterance. Given this lack of specificity, the stimulus argument was analysed as abstract, even though the element to be supplied could theoretically also be concrete. In the present context, the implied verbal action would probably

be equally analysed as abstract, but different expressions of underspecified stimuli and different contexts may limit the possible interpretations further. In the example in (121), the stimulus argument itself contains the predicate *lykede* within a restrictive relative clause:

- (121) for they deden her owne wil in all þyng þat **hem** lykede & hadde of power [...] ‘for they did their own will in all things that pleased them and [which they] had power over’ (CMBRUT3,2.27)

Even though *all þyng* ‘all things’ are defined by the relative clause and at some level specific, their precise nature remains unspecified, and the stimulus argument in this example and in similar observations was hence analysed as abstract.

Other stimulus arguments depend on the context for their interpretation as either concrete or abstract. The word *company*, for example, whose synonym *felawshep* was already encountered in (116) above, can denote either a concrete group of people or the abstract notion of another person’s presence. The latter sense is implied in (122):

- (122) she is soe much a woman , that I am almost ashamed to say I am her Aunte , and soe Pritty that if I had any designe to gaine a Servant I should not like her company . (OSBORNE,52.024.1160)

Another case in point are verbal nouns, which were generally interpreted as abstract, but which, in some cases, may also designate concrete things like the result of a verbal action. Consider the example given in (123):

- (123) and I desire to know wheather **he** likes the meath , and wheather my brother had the pyes [i.e. pies] I sent him . (HARLEY,149.042.1267)

In this passage, the stimulus argument *the meath* could refer to the general action of mowing, but also to the product of that action, viz. the mowed crop. Since the letter does not provide any more context, the former interpretation was decided upon as the more likely one and the instance classified as abstract.

One fairly frequent type of stimulus argument in the *PCEEC* whose interpretation can be ambiguous involves acts of communication, which can be thought of as abstract verbal actions or as concrete instantiations of these actions, either in written or in spoken form. An admonition, for example, constitutes a verbal action, but the context of the passage cited in (124) makes it clear that a written document is referred to:

(124) I do lyke notably well the cōpy of your admonition , which can not be ill taken of any that hath the spirit of unity [...] (FITZHER,13.002.32)

Such written documents, but also the oral delivery of a specific text in the form of a speech, were analysed as concrete stimuli, just like the more straightforward examples of letters and other physical objects. Further examples include *your answere* (PARKHUR,204.054.998) and *my poore advice* (PARKHUR,246.083.1479), but their interpretation is not disambiguated by the respective contexts, and they were therefore interpreted as verbal actions by default and categorised as abstract stimuli.

3.4.2.3. Category

While the two variables animacy and abstractness were intended to capture the semantic properties of the participants that occur as arguments of impersonal verbs and, by extension, their potential to be interpreted as agentive participants, the realisation of these participants by different lexical categories such as pronouns and nouns or by different syntactic categories such as NPs and clausal complements is a grammatical feature that applies to the level of linguistic expression rather than linguistic content. Its inclusion as a potential variable is of interest for two main reasons. First, the contrast between pronominal and nominal constituents is one of the conditions originally proposed by Timberlake in his treatment of the actualisation of genitive subjects in Finnish participial clauses (cf. Timberlake 1977: 153–156 and the exposition in section 2.2.2.1.1 above), and it seems reasonable to assume that such a distinction can also be relevant for the changes in the construction type of experiencer verbs in English, since these involve the case marking of nominative-subject constituents. Crucial to the present discussion is the fact that the categorial distinction takes part in a hierarchy involving different levels of agency (cf. Timberlake 1977: 156), which are given in a slightly adapted form in (58) above, repeated here as (125):

(125) pronouns > agentive nouns > non-agentive nouns

According to this hierarchy, pronouns adopt the innovative construction earlier than full NPs, and agentive nouns adopt it earlier than non-agentive nouns because the former are considered to be more subject-like than the latter. While this distinction is easily transferrable to the present investigation, it is worth noting that the above hierarchy is inferred from a rather small set of empirical observations in Timberlake's contribution. Its validity thus requires further motivation by theoretical considerations. One argument in favour of a categorial distinction is

the fact that personal pronouns, to which the term pronouns in the above hierarchy seems to apply first and foremost judging by the examples given, that such pronouns denote concrete and animate participants, while the referents of full NPs, even in the case of agentive nouns, are potentially less concrete and not necessarily always conceptualised as animate. The hierarchy of lexical (and syntactic) categories thus appears to be motivated, to some extent at least, by the semantic properties animacy and abstractness, which were discussed in the previous sections, and this would make the inclusion of category as a separate variable potentially redundant.

A problem with the application of category as a variable based on semantic properties is the fact that these are not normally consistent across a single categorial realisation. While personal pronouns are relatively unified in their reference as a result of their definition, other categories such as NPs or PPs are not limited in the same way, unless specific subcategories such as agentive and non-agentive nouns are distinguished. For PPs, it is, in fact, the complement of the preposition that is relevant for the semantic properties of the referent. This could lead to a further distinction of PPs with a pronominal complement and PPs with a nominal complement, which could either form different subgroups on their own or merge with the respective simple categories of pronouns and full NPs. Either way, the semantic properties of the referents would not be fully captured. Another example are null constituents, which contrast with other constituents as a distinct syntactic category, but whose referents are not unified in any specific way, even if they are treated as equivalent to pronominal constituents, which can be optionally deleted under coordination in English and are optional to begin with in so-called pro-drop languages. This inconsistency in the referential properties of lexical and syntactic categories makes it difficult to use semantic properties of their referents as a theoretical motivation for the impact of such categories on the diachronic changes under discussion. At the same time, a distinction of different categories according to the semantic properties of their referents, for example one between pronouns with animate referents and pronouns with inanimate referents, would yield the variable category redundant, since, ultimately, the factor to be tested would be the semantic properties, which are already tested for in a more direct way by the respective variables.

The problem is particularly apparent in the distinction of different pronominal categories. Personal pronouns, on the one hand, are relatively homogeneous in their reference, since they tend to refer to actual persons, while the same is not true of anaphoric and relative pronouns, for example, which have endophoric reference and thus do not immediately identify their referents in the context of a given discourse. While personal pronouns refer fairly consistently to concrete, animate referents, anaphoric pronouns such as *they* can refer to either

animate or inanimate referents, and anaphoric *it* can even refer to propositional arguments, which are regarded as the most abstract type of participant and certainly more abstract than the discourse participants represented by first- and second-person pronouns. Furthermore, it is not clear which or how many pronominal subcategories, if any, should be distinguished within a pronominal category. Fine-grained distinctions such as those between indefinite pronouns, interrogative pronouns and demonstrative pronouns, all of which occur as arguments of the impersonal verbs under discussion, would certainly be possible conceptually, but their significance for the dependent variable construction type, both in terms of its theoretical motivation and in terms of its detection by sufficient quantitative representation, is largely negligible. The least problematic decision thus seems to be a more general distinction between personal pronouns, which have relatively homogeneous exophoric reference, on the one hand and the remaining pronominal items, which have heterogeneous endophoric reference, on the other.

An additional point that needs to be addressed in this context is the definition of personal pronouns, since there are certain differences in the terminological conventions between different linguistic disciplines. For Present-Day English, personal pronouns are usually taken to include pronouns of the first and second person as well as third-person pronouns. By a narrower definition, personal pronouns are, instead, limited to the first and second person, since these most commonly refer to actual persons as the primary discourse participants, while third-person pronouns can be classified as anaphoric pronouns, whose resumptive function applies to any kind of antecedent. The wider definition for Present-Day English is understandable in light of the fact that certain English pronouns of the third person are also predominantly used for personal referents. This is a result of the fact that English has undergone a transition from a gender system based on grammatical gender to a system based on natural gender. The equivalents of PDE *he* and *she* could still be used anaphorically for any kind of referent with masculine or feminine grammatical gender until the EME period, but they are now restricted to masculine and feminine referents in terms of natural gender, viz. to human referents or, occasionally, animals, and certain other exceptions. This means that, in a diachronic scenario, neither the wider nor the narrower definition are able to capture the referential properties of third-person pronouns, since different systems apply at different times. The solution adopted here is to include the aforementioned third-person singular masculine and feminine pronouns in the definition of personal pronouns and to exclude other third-person pronouns, viz. *it* and *they* and their ME equivalents, which do not refer consistently to persons in either Early Modern English or Present-Day English. This definition does, of course, partially correlate with the

semantic properties in terms of animacy and abstractness, which means that it appears to be tailored towards criteria that lie outside the variable of lexical and syntactic category directly. On the other hand, this is only partly true, since the difference between personal and non-personal pronouns according to the above definition corresponds more generally to a distinction between homogeneous exophoric reference and heterogeneous endophoric reference. In either case, it opens up the opportunity of investigating the effect of personal pronouns the way they were presumably conceptualised in Timberlake's original proposal.

As it stands, the categorial realisation of the arguments of impersonal verbs represents a syntactic variable in its own right and is intended to be independent of the semantic features of its referents. The levels that were ultimately distinguished include personal pronouns as they were defined above, all other pronouns as a second pronominal category of non-personal pronouns, null constituents, NPs, PPs and clausal constituents, the latter including both finite and non-finite subordinate clauses as a single category. An exception to the above classification of pronouns are reflexive pronouns such as *myself* and *herself*, which were included in the category of personal pronouns. This is because they are restricted in terms of their reference in the same way as personal pronouns, and because their use in the data was generally of an emphatic rather than a reflexive nature, which eliminates the categorial contrast for the present purpose. Given the potential of most of the distinguished categories to refer to a wider range of referents with different semantic properties, any measurable effect of the variable category on the construction type used with impersonal verbs can be assumed to independently reflect a syntactic feature, which can function as a control variable besides the two semantic variables discussed in the previous sections. If the wide range of referents covered by the different categories leads to a reduction of the variable's predictive power, this would be adequate in reflecting that the categorial distinction itself is not an essential factor in the diachronic development of impersonal verbs. As with animacy and abstractness, some examples of both experiencer and stimulus arguments will serve to illustrate the different categorial realisations that occur with these constituents and also to exemplify the syntactic categories that have not been fully discussed in the present section.

3.4.2.3.1. Experiencer

The different categorial realisations of the experiencer argument were recorded for all observations of impersonal verbs that were retrieved from the corpora, but any effect they might have on the construction type cannot be meaningfully investigated for the diachronic period

that is of interest here, since the ability to bear overt case, which is the most important criterion for the construction type, is generally not present in NPs with nominal heads but restricted to certain types of pronominal constituents during the periods of English under investigation. The contrast between pronominal and nominal experiencer arguments will thus be present only in the subset of instances in which the construction type was established by other criteria, viz. the occurrence in certain constituent order types with an expletive, the case-identity constraint or verb agreement, while those instances that exhibit the central criterion of case marking will be largely restricted to pronominal items. Observations which remained ambiguous in terms of their construction type due to a lack of applicable criteria were excluded from the statistical analysis, and this implies that nominal constituents are underrepresented in the final dataset compared to their total frequency in data initially extracted. Since the composition of that dataset is thus partly dependent upon the categorial realisation of the experiencer argument, this feature cannot be used as a predictor variable. An overview of the different realisations of the experiencer argument will be interesting nevertheless, since it can provide the basis for an investigation of the claim that experiencer arguments have a particular tendency towards realisation by personal pronouns (cf. e.g. McCawley 1976: 196). This would make structural ambiguity an unlikely starting point for the changes in the case marking of the experiencer argument, since personal pronouns typically retain their distinctive case forms in English. The relative frequency of ambiguously and unambiguously case-marked categories is thus an additional point of interest in recording the categorial realisation of the experiencer argument.

The range of lexical and syntactic categories that realise the experiencer argument generally excludes clausal constituents, since these refer to the propositions, which do not qualify as experiencing entities. All of the other categories distinguished above occur, however, including personal as well as non-personal pronouns, null constituents, NPs and PPs. An example of an experiencer argument realised by a personal pronoun is given in (126):

(126) Among an heap of sterres fixe it liked **me** for to take the altitude of the faire white sterre that is clepid Alhabor ,

‘Among a heap of fixed stars, I pleased me to measure the altitude of the fair white star that is called Alhabor’ (CMASTRO,670.C2.240)

While the pronominal experiencer in the above example occurs in its distinctive oblique case form *me*, the non-personal pronouns do not necessarily show this kind of contrast. Nevertheless, the majority of all pronominal experiencers taken together are, indeed, case-marked, since non-personal pronouns are much less frequent than personal pronouns in this function, and since

their most frequent representatives, viz. the third-person plural pronoun *they* and its variant *heo*, which were categorised as anaphoric pronouns, even though they typically denote personal referents in the context of experiencer arguments, equally maintain at least two distinct case forms during the ME period. Other non-personal pronouns include the indefinite pronoun *one*, the identity pronoun *the same* and the relative pronouns *that*, *which* and *who*, the last of which potentially contrasts with an oblique form *whom*. Two examples are given in (127) and (128):

(127) I shewed and communicated al thing to my Lord of Norfolke , **who** liked al thing very wel . (GARDIN,30.044.161)

(128) [...] his Highnes therefore hartely requyreth your Grace , that it may lyke **the same** to send vn to hym the said instructions that his Grace may by the same be lerned of your Gracis prudent advice and counsaile in the premissis .

‘his Highness therefore kindly asks your Grace that it may like the same to send to him the said instructions so that his Grace may by the same be instructed of your Grace’s prudent advice and council in the aforementioned matters’ (MORE,296.016.209)

The relative pronoun *who* in (127) does, in fact, contrast with an oblique form in the letter from which it is taken and thus indicates a personal construction with a nominative experiencer. One could argue that this specific relative pronoun would need to be classified as a personal pronoun according to the definition developed above, since it is typically restricted to animate referents in Present-Day English, but this restriction and the associated contrast with non-personal *which* only emerge gradually in later Middle English, which is also evidenced by the fact that the latter pronoun equally occurs in reference to an animate participant in the corpus. Similar to *which*, the identity pronoun *the same* can refer to either animate or inanimate participants, and both cases are, in fact, illustrated in the text token in (128). Its function with the animate antecedent seems to be to maintain a third-person address instead of the personal pronoun *you* that would otherwise be appropriate in this context. Since such specific differences among non-personal pronouns can hardly be measured in terms of their effect on the construction type, they were simply subsumed as a single category.

Apart from pronouns, experiencer arguments are realised by null constituents, full NPs and PPs. The last of these were regarded as equivalent to constituents with oblique case marking, and the relevant experiencer arguments are thus consistently analysed as being part of impersonal constructions. Since the syntactic category contributes to the definition of one of the levels of construction type in this case, it is obviously not a suitable predictor. Still, it is

interesting to note that the difference between experiencer and stimulus is also brought out in the use of different prepositions for these arguments. For the experiencer, these primarily include *to*, *unto* and *till* in finite active clauses and *by* and *of* in passive clauses or participial clauses. It is also worth noting that the realisation of experiencer arguments by PPs is particularly frequent in certain periods. With *liken*, this realisation is, in fact, the most frequent one in M3 (1250–1350). Examples of NP and PP experiencers are given in (129) and (130):

(129) Ðan þe Erchebisshop likyd wel þe tale

‘then the archbishop liked the tale well’ (CMKEMPE,127.2953)

(130) For parauenture , whan it likip vnto God , þoo þat mowe not at þe first tyme haue it [...] bot seeldom & þat not wiþ-outen grete trauayle , siþen after þei schulen haue it whan þei wile , as ofte as hem likip .

‘For perchance, when it pleases God, they who may not have it [sc. perfect contemplation] at first but seldom and that not without great effort, later they shall have it whenever they want, as often as they please’ (CMCLOUD,128.749)

The PP experiencer *vnto God* in (130) appears to be due to Anglo-Norman influence, which is evidenced by the French-derived diction *parauenture* and *trauayle*. The example may point to a more general dependence of PP experiencers on Anglo-Norman influence (cf. Trips/Stein 2019), but this was not investigated further here. The NP experiencer *þe Erchebisshop* in (129) is typical in that it designates a social function or, more specifically in this case, a member of the cleric. Similar instances include functions like that of bishop or king, but also kinship terms and proper nouns occur as realisation of the experiencer. The most frequent instances of NP experiencers, however, are those of abstract reference such as *your Grace* and *your Highness*, which are widely used for the purpose of formal address or indirect reference. Although NP experiencers are thus represented more widely in the data, their typical context is not that of narrative texts but that of formal correspondence.

Finally, experiencer arguments, just like other topical constituents, are often deleted under coordination and thus occur as null constituents without phonological or morphological form. As such, they cannot exhibit case marking, and the observations in which they occur are frequently not classifiable in terms of their construction type, unless they can be disambiguated by one of the other criteria used for this purpose. The antecedent of deleted constituents was generally recorded in order to ascertain the semantic properties of the referent, but since the antecedent of null constituents is often pronominal itself, extensive recourse to the context was

a requirement in some cases. It should also be noted that the morphological form of an antecedent cannot be used in order to determine the covert case marking of the deleted constituent, since the case marking of the former is not necessarily representative of the deleted constituent. No assumptions about the covert case marking of null constituents were therefore made, especially not for the ME period of transition. An example of a deleted experiencer argument is given in (131):

- (131) I_i have read the copy of your letter and ____i lyke it singularly well .
(FITZHER,4.001.3–4)

The immediate antecedent of the deleted constituent is I , which can be identified by the wider context as Thomas Fitzherbert, the author of the letter from which the instance is taken. The semantic properties of the referent are easily established as human and concrete, which is, of course, rather trivial given that the antecedent of the experiencer argument takes the form of a personal pronoun in the present case. The construction type can be established based on the agreement of the verb form *lyke*, which reflects the grammatical properties of the deleted experiencer argument, while the stimulus argument *it* would require morphological marking of the third person singular in the indicative mood. As mentioned above, the usefulness of verb agreement for determining the construction type of impersonal verbs was generally rather limited, since two arguments with identical grammatical properties cannot be distinguished as controllers of verb agreement, and since verbs in the subjunctive mood or the past tense do not exhibit distinctive person marking. The realisation of the experiencer argument by a null constituent is thus often an obstacle to the interpretation of the dependent variable, which leads to a certain increase in the proportion of ambiguous contexts. For stimulus arguments, the presence of overt case marking is not as crucial to the analysis as it is with experiencer arguments, since the occurrence of either oblique or nominative stimuli does not affect the definition of impersonal constructions developed here. The categories by which stimulus arguments are typically realised in the two corpora underlying this investigation will be outlined in the following section.

3.4.2.3.2. Stimulus

The stimulus argument is generally realised by the same categories as the experiencer argument, but, in addition, stimulus arguments can take the form of finite and non-finite clauses, which are collectively referred to as clausal constituents or propositional arguments in the present

context. For a verb like *liken*, such clausal constituents are, in fact, the most frequently attested categorial realisation, while the proportion of personal pronouns is lower and the proportion of non-personal pronouns higher compared to that observed for experiencer arguments. The subcategories of the latter are also more varied and include interrogative, demonstrative and possessive pronouns in addition to the subcategories discussed above. The majority of instances of non-personal pronouns take the form of anaphoric pronouns or relative pronouns, however, the former of which specifically include the third-person singular pronoun *it* and the third-person plural pronoun *they* and their ME equivalents. Their use is exemplified in (132) and (133) below. The other pronominal subcategories occur more sporadically and include the demonstrative pronouns *this* and *that*, the identity pronoun *the same*, indefinite pronouns, interrogative pronouns and possessive pronouns. The passage in (134) contains an example of the use of a relative pronoun:

(132) and for ði ðu wendest þat ic ðe ware ilich , þat hit likede **me** swa swa hit dede ðe

‘and therefore you believed that I was like you, that it pleased me so as it did you’
(CMVICES1,11.110)

(133) I doe earnestly desir to know how **yo=r= Lo.** [i.e. your Lordship] doth like them , that if any thing be distastefull unto y=u= [i.e. you], it may be amended if God send me life to the next yeere , (ARUNDEL,38.002.13)

(134) and than every knyght toke the way that **hym** lyked beste .

‘and then every knight took the way that pleased him best’ (CMMALORY,638.3880)

The first example in (132) contains an instance of the anaphoric pronoun *hit*. The constituent order is comparable to the frequent *it_V_E* order in constructions with an expletive, but there is no extraposed clausal constituent whose position the pronoun is intended to fill. Instead, the use of the pronoun has anaphoric function and refers back to a propositional argument in the preceding context, whose substance can be paraphrased as ‘before him (sc. your brother) you spoke good, behind him evil so that he did not hear’. An alternative analysis would be to regard the pronoun as an instance of the expletive in conjunction with an inferred stimulus. Since the propositional argument to which *hit* refers does not immediately follow the clause that contains *hit*, the proposition needs to be recovered by inference one way or another, and the main difference between the two analyses is thus whether *hit* is assumed to be anaphoric or cataphoric. The former option involving the analysis of *hit* as a referential pronoun was preferred here. The second example in (133) contains the oblique form of the anaphoric pronoun

they, which refers to a non-human, albeit animate, participant, viz. one larger and some smaller fish, which the author of the letter presumably sent as a present to her kin. It is most likely that the fish are dead at the time of writing the letter, and they thus pattern semantically with inanimate participants. The relative pronoun *that*, which occurs as stimulus argument in (134), equally refers back to an inanimate participant and thus contrasts with the personal pronoun *hym*, which realises the experiencer argument in the same instance.

Stimulus arguments can also be realised by personal pronouns, although this occurs less frequently with most verbs. Two examples are given in (135) and (136):

(135) for hie wule liken alle þe lechures þe on hire lokeð . and swo dragen hem to hire .

‘for she wants to please all the lechers that look at her and thus draw them to her’
(CMTRINIT,29.386)

(136) but otherwise I can assure you I have heard from People that prettend to know her very well , that her kindenesse to Compton was very moderate , and that **she** never liked him soe well , as when hee dyed and gave her his Estate . (OSBORNE,53.024.1178)

In the first example in (135), the personal pronoun *hie* contrasts with an experiencer expressed by a full NP, which is the opposite of the more common distribution. In the second example in (136), both arguments are realised by personal pronouns, the experiencer argument *she* and the stimulus argument *him*. Since, in both instances, both participants are animate and thus capable of undergoing experiences, the distribution of the semantic roles needs to be established by the context. This is easily possible within the extent of the citations above. In the first example, the purposeful nature of the act of pleasing is apparent from the coordinated clause *and swo dragen hem to hire* ‘and thus draw them to herself’ as well as from the volitional auxiliary *wule* ‘wants’, both of which imply that the controlling constituent *hie* is the stimulus. In the second example, a glossing of *liked* with ‘pleased’ would be theoretically possible, but since the predication obtains after the decease of the second participant, the first participant can be unambiguously identified as the experiencer. Strictly speaking, the second participant could also be analysed as inanimate in this context, as a similar argument was made about the fish in (133), but since the conceptualisation of the referent as human probably still applies, it was analysed the same way as human referents elsewhere. Overall, the distribution of the different types of pronouns for stimulus and experiencer arguments is not very surprising and largely correlates with the different distribution of human and non-human referents.

The PPs that encode the stimulus argument of *liken* involve a variety of prepositional heads, but only one of them is of considerable frequency, while the other prepositions are found only sporadically. One such exceptional case is the use of *on*, which occurs in a translated citation from the Gospel of Matthew in *Vices and Virtues* but not elsewhere in the present data. In its use of the preposition, the translation appears to be dependent on the Latin exemplar, which is cited immediately preceding the instance, but it is also not an exact glossing. The Latin version and the ME rendition are given below in (137) and (138) respectively:

(137) [...] Hic est filius meus dilectus , in quo **mihi** bene complacui ,

this is son my dear, in whom me-DAT well pleased-1-SG-IND-PF

‘this is my dear son, in whom I have pleased myself well’ (CMVICES1,119.1476)

(138) ' Ðis is mi leue sune , on him **me** likeð swiðe . '

“‘This is my dear son, in him I am well pleased’” (CMVICES1,119.1477–1478)

While the Latin original is constructed personally with a first-person subject and a coreferential experiencer argument *mihi* ‘myself’, the ME translation uses an impersonal construction with a third-person singular verb form, which eliminates the reflexive relation between stimulus and experiencer and promotes the element in regard to which the subject is pleased to the role of stimulus. Another preposition that is used infrequently with stimulus arguments is *wiþ* ‘with’. It only occurs once in a collection of vernacular sermons called the *English Wycliffite Sermons*, which do not have an immediate Latin exemplar. Instead, the use of the preposition with an argument of *liken* cited in (139) below is reminiscent of the cognate Icelandic verb *lika* ‘to like’, which is quite commonly constructed with the equivalent preposition *við* ‘with’. Alternatively, the verb form might also derive from ME *lukken* ‘to happen (fortunately)’, which would be comparable to the verb *gelükken* ‘to succeed’ in Middle Low German:

(139) And so he wente to heþene folc þat weren wiþowte grace , and 3et **hym** lykede not wiþ hem for þer kyndely resoun .

‘And so he (s.c. the fiend) went to the heathen people, who were without grace, but still he did not like them because of their natural reason.’ (CMWYCSER,410.3319–3320)

If interpreted correctly as arguments of *liken* ‘to please’, the two instances in (138) and (139) yield the closest equivalent in the data to “true” impersonal constructions, i.e. type-i constructions without a nominative constituent. This is not the inherited construction of the type-ii verb *liken*, however, which means that the prepositional stimuli do not immediately

replace constituents with oblique case marking. This is also indicated by the fact that other instances of prepositional stimuli mainly occur in much later periods. An example of a preposition that occurs sporadically in late Middle English is *by*, which, apart from one exception, is used exclusively by different authors of the *Paston Letters* (1425–1519?), cf. the example in (140) below:

(140) and he seyde to here **he lyked** wel by þe chere þat I made hym .

‘and he said to her he liked the chair that I made him well’ (PASTON,I,30.008.124)

The virtual restriction to this particular collection of letters may point to a geographically limited distribution of *by* as a preposition introducing stimulus arguments of the verb *liken*, but it could also simply be part of the idiolect of the correspondents involved.

In contrast to the singular occurrences of *on* and *with* and the restricted occurrence of *by*, the preposition *of* occurs fairly widely with stimulus arguments of *liken*, albeit during a rather limited time period. The following example in (141) is taken from a letter to the Earl of Leicester, which is dated to 1585:

(141) But according as **her majestie** shall lyke of the proceading with the cause , so must the advise be .

‘But according to how her Majesty shall like the proceeding with the cause, so must the advice be.’ (LEYCEST,418.088.2720)

Similar examples occur in other letters from E2 (1570–1640) and, less frequently, from the earlier half of E3 (1640–1710). In some cases, the preposition occurs “stranded” at the end of a relative or comparative clause, or even in absolute constructions without an expressed stimulus argument. Examples of each construction are given in (142) and (143):

(142) I have sent unto you Goodman Cooper , one hoome [i.e. whom] I thinke **you** will well like of , (OXINDE,30.013.128)

(143) My very good lord , I shuld be ashamed greatly for not oftenar wrytyng to your lordship of late , having receaved so many [sc. letters] from yow , but that I have an excuse more sufficient than **I lyke** of , which also this beror can inform yow of .

The grammatically unexpressed stimulus in (143) can be inferred as *having* (i.e. *an excuse more sufficient than I lyke of [having]*), and the persistent use of the preposition even in cases where the object of the preposition does not immediately follow or is absent altogether shows that the

association of *liken* with *of* is a particularly strong one in this period. Its frequency drops during the subsequent period, and its latest occurrence in the *PCEEC* was found in a letter dated to 1681. While it would be interesting to establish the exact distribution of the preposition and to ascertain whether the use of prepositional complements with transitive verbs is part of a wider phenomenon, this is beyond the scope of the present investigation.

Stimulus arguments are occasionally also realised by null constituents just like experiencer arguments, but, unlike the latter, these instances rarely affect the analysis of the construction type, which is primarily based on the case of the experiencer. An example of a construction that remains ambiguous is given in (144):

(144) For ði ne mai wexen non god sad of gades wordes on zeure herte molde , for ðan michele embeðanc ðe zie habbeð on zeure michele wele , ðe zie michel zitsið , and luuieð and likeð , and draheð to forliesen .

‘Therefore, no good seed of God’s words can grow on the mould of your hearts, because of the great concern that you have for your great riches, which you covet, love and dread to lose.’ (CMVICES1,69.777)

In this example, both the experiencer *zie* ‘ye’ and the stimulus *ðe* ‘that’ are deleted in what is essentially an enumeration of verbs, which means that no classification of the construction type based on overt morphological features is possible. Although the experiencer and the stimulus have contrasting grammatical properties, represented by the second-person plural pronoun and the third-person singular collective noun *wele* ‘riches’, this contrast is not distinctively marked on the verb, which means that no classification of the construction type based on agreement properties is possible either. The effect that the realisation of the stimulus argument by null constituents is expected to have on the construction type is difficult to motivate theoretically, although null constituents are probably best compared to pronominal constituents, since both of these categories refer to topical referents. Since null constituents may be regarded as the equivalent of either personal pronouns or non-personal pronouns, they could be identified with either of the two pronominal categories. This could be done in accordance with the semantic properties of their referents, but since the variable category is intended to reflect a syntactic distinction and not primarily animacy, which is, of course, similarly problematic for the distinction of the two pronominal categories to begin with, null constituents were regarded as a separate category. The problem will be taken up again in section 4.1.3.1.2.1, which describes the configuration of the variable category for the regression model.

The final category to be illustrated here involves clausal constituents. As mentioned above, finite and non-finite subordinate clauses were subsumed under a single category and are referred to by a superordinate term as propositions. This term reflects the semantic equivalence of finite and non-finite clauses as complex statements, but since the categorial realisation of arguments is intended to function as a syntactic variable and not as a semantic variable, and since different clause types may have different associations in the data, finite and non-finite clauses were also recorded as separate categories in order to facilitate future reference on a more specific level. Their conflation into a single variable level, on the other hand, was partly due to pragmatic considerations, since propositional stimulus arguments were, in many cases, inferred from the context and did not exhibit an overt grammatical form. The inference of their form as either finite or non-finite clauses was primarily based on the criterion of subject identity, since the difference between finite and non-finite subordinate clauses is often one between non-identity and identity of the subject of the subordinate clause and a topical element of the superordinate clause. The resulting classifications were generally undertaken with great confidence. Since the inference of finite and non-finite clauses nevertheless involves some ambiguity, the more general level of propositions was chosen as variable level as a measure of caution against the few uncertain instances that remained.

An example of an inferred infinitival stimulus is given in (145), which is taken from the twelfth-century exegetical poem *Ormulum*:

(145) | Gast ȝifeþþ herrte , & wille , & mahht , || Þær þær **himm sellfenn likeþþ** , |

‘The Holy Ghost gives heart and will and might where it pleases him’
(CMORM,II,226.2419)

The stimulus in this example can be rendered as PDE *to give heart and will and might*, but it is not grammatically expressed in the ME text and thus needs to be inferred from the preceding context. Its implied subject is coreferential with the experiencer argument *himm sellfeen*, which refers back to *Gast* ‘ghost’, and its most economic expression is therefore by a non-finite or, more specifically, by an infinitival clause. This type of clause is also the more frequently inferred type compared to finite subordinate clauses. Although the inference of a similar expression with a finite verb form *that he give(s) heart and will and might* would be possible in the present case as well, this would rather be expected in the case of non-coreferential constituents. An example of such a case of non-coreferential constituents is given in (146):

(146) But 3if it lyke 3ou I schall schewe how 3ee schull knowe & preue to the ende þat 3ee schull not ben disceyued .

‘But if it should like you, I shall demonstrate how you shall know and verify to the end that you shall not be deceived.’ (CMMANDEV,33.825)

In this example, the stimulus argument was inferred in the form of a finite subordinate clause that would translate to PDE *that I show you how you shall know* etc. The establishment of its grammatical form is based on the fact that there is no identity between the subject *I* of the inferred proposition and the topical experiencer argument *3ou* ‘you’ in the clause *[b]ut 3if it lyke 3ou*, in which the clausal stimulus would be embedded. A passive infinitival clause *to be shown*, which could otherwise be invoked as complement of the same clause, would only include part of the semantic proposition, omitting the subject *I*. The two examples in (145) and (146) thus illustrate, once again, the general problem of classifying logical arguments that are grammatically unexpressed in terms of their syntactic category. Since these are not really deleted instances of a specific antecedent but rather semantic arguments implied by the surrounding context, they were not identified with the null constituents discussed above but analysed in terms of the syntactic constituents that would be equivalent to the semantic propositions. These points are important because the impact of inferred clausal stimuli on the overall composition of the members of the category of propositions can be quite considerable. In the dataset containing finite active clauses with two arguments of *liken* (n = 745), for example, their proportion is about 22.7% of all propositional stimuli, which means that almost one fourth of all clausal constituents are merely inferred from the context.

Apart from inferred clausal constituents, there are, of course, numerous examples of grammatically expressed stimuli that take the form of finite or non-finite clauses, either with or without an expletive. Since these represent predefined categories, they do not require much theoretical discussion and are mainly exemplified for illustrative purposes and for the sake of completeness. A special type are instances embedded in relative clauses, in which the fronted relativising element precedes the verbal predicate of the non-finite clause. An example of this case is given in (147):

(147) Thanne seyden they with o [i.e. one] voys , " Worshipful lady, we putten us and oure goodes al fully in youre wil and disposicioun , and been redy to comen , what day that it like unto youre noblesse to lymyte us or assigne us , for to maken oure obligacioun and boond as strong as it liketh unto youre goodnesse , that we mowe fulfille the wille of yow and of my lord Melibee . " (CMCTMELI,237.C1.784)

The infinitival complement *what day ... to lymyte us or assigne us* is separated by the predicate *liken* and the experiencer argument *unto youre noblesse*, since the object of the non-finite verb form *lymyte* is fronted together with the relativising element *what*. The whole proposition is anticipated by the expletive constituent *it*. Similar instances occur in the *PCEEC*, an example of which is cited in (148):

(148) Jon Hobbys tellith me þat ye be seekly , wech **me** lekith not to here ,

‘Jon Hobby tells me that you are sickly, which I do not like to hear.’
(PASTON,I,134.036.790)

In this case, the clausal stimulus *wech ... to here* is not anticipated by an expletive, perhaps because the preposed internal object *wech* can be initially parsed as the subject of the relative clause, which is later expanded by the non-finite verb form. Incidentally, the relative pronoun itself refers back to the clausal constituent *þat ye be seekly*, rendering the stimulus argument rather complex. An example of an expressed finite subordinate clause is given in (149):

(149) botte 3it lykes **hym** bettyr that we take fullye his blessedde blode to wasche vs with of synne ,

‘but he likes it better that we fully take his blessed blood to wash us with of sin’
(CMJULNOR,50.3)

Again, the topical experiencer argument *hym* and the subject of the propositional stimulus argument *we* are not coreferential, which motivates the overt form of the latter in combination with a finite verb form. Finally, the most common type of clausal stimuli are infinitival complements, which occur even more frequently in explicitly expressed form than they do in inferred form with the verb *liken*. A typical example is given in (150):

(150) And 3if **3ou** lyke to here how the mele cometh out of the trees I schall seye 3ou .

‘And if you would like to hear how the flour comes from the trees, I shall tell you.’
(CMMANDEV,126.3063)

The construction of this example is superficially similar to Present-Day English, but, judging by the evidence of further instances in the surrounding context of the passage, the pronominal experiencer *3ou* presents a contrastive oblique case form in what must hence be an impersonal construction with a verb in the subjunctive mood.

3.4.3. Diachronicity

Apart from the conceptual variables, whose impact on the construction type the present study is designed to test, the logistic regression model needs to account for the fact that the diachronic data involve a near-complete shift in the odds of impersonal and personal constructions for most of the verbs under discussion. Without the implementation of a variable that represents this fact, an uneven diachronic distribution of the levels of a predictor variable would lead to potentially spurious associations with the dependent variable, since a high frequency of a given variable level during one of the later periods, for example, would be interpreted as an association of that level with personal constructions, whose odds continuously increase over time. The temporal dimension thus presents an additional factor that needs to be controlled for in the model. From the point of view of language production, each observation of a given verb does, of course, represent a synchronic choice between the options available at a given point in time, but such a level of conceived synchronicity cannot be adequately accounted for by a diachronic corpus that lacks the sufficient frequencies to represent the relevant variables within each of the diachronic layers distinguished.

The most natural way to control for time in a regression model of the diachronic change in the odds of the two construction types would be by implementation of a continuous variable, which could be used to measure time in units of varying size. Of course, such a variable only serves as proxy for the actual causative forces underlying linguistic change in any scenario, but the present case poses several more specific problems to the representation of time on a continuous scale. First of all, the periodisation of English, which presents the most readily available type of chronological information contained in both corpora, is not very fine-grained. The measurement of time in units as large as centuries may, of course, be acceptable in investigations of large-scale diachronic changes, but the period divisions in the *PPCME2* and the *PCEEC*, which follow the standard divisions of the *Helsinki Corpus*, are also of varying length, which means that they are not immediately equivalent to units of time. An overview of the relevant periods and their length is given in table 4 below:

M1	M2	M3	M4	E1	E2	E3
1150-1250	1250-1350	1350-1420	1420-1500	1500-1570	1570-1640	1640-1710
100 years	100 years	70 years	80 years	70 years	70 years	70 years

As can be seen, the periods span across 80 years on average, with the earliest periods M1 and M2 being somewhat longer and the EModE periods E1–E3 and M3 being somewhat shorter than average. Despite this difference in length, the period information could, perhaps, still be implemented as a continuous variable, which would represent time in units of roughly 80 years. Alternatively, the philological information that is distributed alongside the *PPCME2*, which contains dates for each text and manuscript in the corpus, could be used to code a separate variable with chronological units of equal size. The precision of this variable would be determined by the level of precision available for the ME manuscripts, which, in some cases, are dated to periods as long as one hundred years, as well as on the optimal combination of these dates into discrete periods. The majority of letters in the *PCEEC*, on the other hand, are datable to the exact year or, in some cases, to the decade in which they were written.

As it stands, not much precision beyond the period information available from the corpora is to be gained by coding a separate variable based on the manuscript dates of the *PPCME2*. In either case, extrapolated predictions of odds based on values on the continuous scale of such a variable would suffer from the imprecise nature of the dates, whether they be measured in units of equal or unequal size. An even more serious problem with the implementation of time as a continuous variable lies in the nature of the data, however. The uneven diachronic distribution of impersonal and personal constructions observable in the data or, more accurately, the uneven changes in this distribution imply that the effect of the elapse of time on the construction type is not a linear phenomenon. This observation is in agreement with the fact that the spread of linguistic change is normally conceived of as an S-shaped process that takes the form of an exponential increase during the initial period and a logarithmic increase during the final period. This variability in the rate of increase of the innovative construction can only be accounted for by treating the different periods as factor levels, which allows for the representation of each period's individual association with the dependent variable, even in cases of non-parametric distribution. Given that the available chronological information is already not very precise for the ME corpus, this downscaling of the diachronic variable seems not very problematic. At the same time, the inclusion of period information as an additional factor in the logistic regression model implies that any variation accounted for by the conceptual predictor variables will be automatically controlled for the variation that the provenance from a specific time period already accounts for.

A number of additional problems associated with the periodisation of the data are of a general nature rather than specific to the conceptualisation of time as a continuous or a categorical variable. One of these is that the data are not distributed evenly among the periods,

a fact which was already discussed in the context of corpus selection. The ME period M2, for example, is represented particularly poorly, both in terms of its general word count and in terms of specific instances of impersonal verbs, while later periods like M4 or E2 are overrepresented by comparison. Another problem, apart from the uneven distribution of the available data, is the fact that the levels of the investigated predictor variables are represented unevenly within the periods as well, which means that the effect of these variables cannot be tested equally well for all periods. Neither of these problems can be easily addressed by additional sampling, since the availability of data from processed historical corpora is limited and cannot be increased at will. The only solution to the uneven diachronic distribution of the data is thus to conflate periods with few observations or with few instances of a given variable level with other periods in order to arrive at larger sample sizes that can be subjected more effectively to statistical inference. Finally, every period division does, of course, involve some arbitrariness, which is less conspicuous in the case of ME texts, whose dates are often approximate to begin with, but which becomes rather glaring in the case of individual letters written during specific years around the time of a period break, some of which fall into the earlier and others into the later period. This problem is difficult to avoid, however, even though identifiable groups of letters could, perhaps, be sorted into periods as clusters that, within a certain margin, include individual instances beyond a given period division. Ultimately, no attempts at smoothing the period boundaries were made in the present study, however, and, instead, the period information was adopted directly from the metadata contained in the corpora.

For ME texts, the *PPCME2* generally gives two types of period information, one relating to the time of composition of a text and the other one relating to the time of production of the manuscript from which a given text or extract is taken. These often coincide, but there are two relevant texts within the corpus, for which the composition date is given as unknown, and several text samples, whose manuscript date falls into a later period than the composition date. The former two texts are the *Trinity Homilies* and the *Lambeth Homilies*, which are thought to derive from the OE period, but whose text samples are based on extant manuscripts from the ME period M1. Other examples include the *English Writings of Richard Rolle*, which receive their *terminus ad quem* from the author's death in 1349 and are thus placed towards the end of the ME period M2, while the relevant manuscript is dated to "a1450", i.e. within a quarter of a century before 1450, and thus falls into the earlier part of M4. This example demonstrates the relative arbitrariness of any periodisation, since the period divisions, in this case, suggest that the manuscript date is removed from the original text by two periods, although the actual difference is probably less than a century and could thus still fall within a single period of the

earlier divisions. Nevertheless, the difference between the text date and the manuscript date remains significant, even if it is somewhat overemphasised by the period divisions, and it requires a decision as to which of the two dates should be used in the diachronic variable. Since there is plentiful evidence that manuscripts of the ME period were copied with the intention of actual use rather than preservation of the original, which is even more plausible in the case of functional texts like the homiletic literature that dominates the ME corpus, the general decision was taken to use manuscripts dates rather than text dates for the temporal classification of the samples. This decision affected around ten different texts, including the ones mentioned above, most of which were thereby dated to M4 rather than M3. Since this number is rather small, a more refined approach of assessing each case individually based on the text type and the actual difference in date would have been conceivable as well, but this was not done, partly for reasons of economy and partly because the limited number of observations from the affected texts that could have been potentially improved was judged to be of low impact. Exceptions to this rule will be discussed for the individual verbs under investigation.

The overall result of the present discussion is a factor with seven period levels, viz. the four ME periods M1–M4 and three EModE periods E1–E3. Since the number of observations from M2 was generally low for all verbs, this period was sometimes conflated with the preceding period M1 for the purpose of quantitative inspection of the proportions. The resulting period of two centuries is, of course, no longer comparable in length to the other periods. This step was necessary, however, since inferences based on small sample sizes are typically not robust, and proportions based on such samples may be misleading. For example, the dataset of *liken* contained one instance of a personal construction in M2, but only four observations within this period in total, suggesting that the proportion of personal constructions had increased to 25% despite the fact that it was much lower again in M3. While this figure is true for the specific sample, it cannot be treated as representative of the construction of impersonal verbs during M2 in general. The quantitative distribution of the observations of *liken* across the different periods ultimately required the conflation of all three early ME periods in order to ensure a statistically robust baseline against which the other periods could be compared. As this early period predominantly exhibits impersonal and only few personal constructions, it functions as a reflection of the initial phase of the theoretically expected S-curve. An equivalent conflation of the two latest periods E2 and E3 achieved a similar representation of the final stage of this process, which, conversely, exhibits mostly personal constructions and only a few remaining impersonal constructions. The conflation of the late periods was not necessary from the point of view of statistical inference, however, since each of the two periods individually contained

a sufficient number of observations in order to display a significant effect on the dependent variable compared to the reference period. The overall design of the period levels is thus, to some extent, data-driven and tailored towards the circumstances of the attestation and the specific development of each verb. The statistical motivation for the case of *liken* will be taken up again in section 4.1.3.1.1.

3.4.4. Additional Properties

A range of additional properties of impersonal and personal constructions were recorded during the coding process but ultimately not employed as predictor variables in the regression model. While some of the recorded properties merely provided useful information for the coding of other variables, others seemed to be potentially relevant as variables themselves, even if they were not included in the present study. They will be briefly described in the present section in order to provide a basis for potential future consideration. The section will begin with the properties of person and number of the two verbal arguments, continue with the semantic property of gender, and finish with additional features of the verbal complex, which include the form of the main verb and the presence of auxiliary verbs.

3.4.4.1. Person

The first additional feature of the experiencer and stimulus argument of impersonal verbs that was recorded is grammatical person. This information was necessary for establishing control of verb agreement in cases where the construction type could not be disambiguated by overt case marking or any of the other criteria discussed above, although the limited distinctiveness of ME verbal forms severely restricted the usefulness of this criterion. Both features are interesting as potential variable of their own, however, since the distinction between discourse participants on the one hand and non-participants on the other as well as that between singular and plural participants potentially affects the interpretation of their agentivity. With regard to person, both arguments are theoretically free to vary despite the common definitional restriction of impersonal constructions to the third person singular of the verb, since this only affects “true” impersonal verbs in type-i constructions. Only the experiencer argument displayed any noteworthy person variation in the present data, however, while the stimulus argument occurred mostly in the third person. This distribution may be partly related to verbal semantics, which can either favour or disfavour the expression of one of the discourse participants as stimulus

argument, but it is also a reflection of the facts already established for the variable animacy, as the person of the stimulus necessarily overlaps with these features.

The experiencer argument shows a markedly different distribution by comparison, despite its general tendency to occur with animate participants realised by personal pronouns. These frequently involve masculine and feminine third-person singular pronouns as well as first- and second-person pronouns, but there are at least two other factors that influence the expression of grammatical person and thus act as potentially underlying predictors. One of these is text type, which, in the present configuration of the corpora, is partly correlated with the dimension of time. Narrative religious texts, which constitute a significant part of the *PPCME2*, are more likely to contain topical experiencer arguments in the third person, although the devotional literature in the same corpus exhibits instances of the first and second person as well, while the second and first person are expected to occur more frequently in the correspondence of the *PCEEC*. The former expectation is borne out by the relevant proportions of experiencer arguments of *liken* illustrated in table 5 below, while the latter expectation is only partly met in that first- and second-person experiencers are more frequent in the *PCEEC* compared to the *PPCME2*, but not or not much more frequent compared to third-person experiencers within the same corpus:

	<i>PPCME2</i>	<i>PCEEC</i>	total
first person	15 (9.4%)	134 (22.9%)	149 (20%)
second person	38 (23.9%)	227 (38.7%)	265 (35.6%)
third person	106 (66.7%)	225 (38.4%)	331 (44.4%)
total	159 (100%)	586 (100%)	745 (100%)

The figures in this table relate to a reduced dataset of *liken* (n = 745), which includes ambiguous besides unambiguous instances of finite active clauses with two arguments, since a further restriction to unambiguous instances would affect the proportion of certain persons due to their correlation with a more frequent realisation by morphologically ambiguous categories. The effect of experiencer person on the dependent variable can, of course, only be measured for unambiguous instances. The table shows that third-person experiencers of *liken* are the most frequent type in the *PPCME2*, where they occur in about 66.7% of all observations, and they are also very frequent in the *PCEEC*, where they occur in about 38.4% of all observations

compared to 38.7% and 22.9% for second-person and first-person experiencers in that corpus respectively. Since observations from the *PPCME2* are generally earlier than those from the *PCEEC*, it can be hypothesised that third-person experiencers have a higher chance of occurring in impersonal constructions, simply because these are more frequent during the earlier periods. On the other hand, this effect should be balanced out by the lower absolute frequency of observations from the *PPCME2*. In order to account for the diachronic nature of the data, the period information thus needs to be entered as a control variable.

A preliminary binomial model of the construction type using the `glm()` function in *R* with experiencer person and period as predictors does, in fact, suggest an increase in the odds of impersonal constructions for third-person experiencers compared to first-person experiencers. However, the same, albeit weaker, effect is shown for experiencers in the second person, which cannot be due to the chronological distribution of the corpora, since second-person experiencers are more than five times as frequent in the *PCEEC*, even after the exclusion of ambiguous instances. The precise details of the model are not given here, since no specific claim regarding the complex development of impersonal constructions is based on it, but it should be noted that the introduction of a second conceptual variable reduces the effect of third-person experiencers on the odds of impersonal constructions significantly, while the effect of second-person experiencers remains virtually identical. The variable in question is the presence or absence of what was discussed in the context of the abstractness of experiencer arguments as indirect reference or formal address, i.e. the use of a possessive pronoun in combination with an abstract quality noun to address or, in some cases, to refer to a concrete human participant. As was stated in that context, the cases of abstract reference are not distributed evenly among the data but concentrated in certain parts of the *PCEEC*, which means that the absence of this factor is not a good predictor of the construction type. Its presence, on the other hand, is highly associated with impersonal constructions in those cases for which the construction type was unambiguously established. In combination with the dominance of third-person experiencers in the earlier *PPCME2*, this probably accounts for the positive impact of such experiencer arguments on the odds of impersonal constructions, but whichever effect is the strongest here, it appears that grammatical person itself is merely a covariant of other factors that afford a more useful explanation. Since an independent effect of experiencer person on the construction type remains doubtful or at least difficult to ascertain within the specific constellation of the present dataset, it was ultimately not included as a predictor variable.

3.4.4.2. Number

The second feature that was recorded for the purpose of establishing verb agreement is the grammatical number of each constituent. With the exception of the dual form *zunnē babe* ‘you both’ in the *Ormulum* (CMORM,I,155.1279), the realisation of this category is restricted to singular and plural forms in the data. These forms do not always refer to semantically singular and plural referents, however. The plural form *ye/you* develops into a polite form and eventually into the standard form of the second person in English, which effectively eliminates the number contrast in that person. Semantically singular participants are, in fact, more frequent as referents of *ye/you* than plural participants, and since the pronominal form itself is rather frequent in the correspondence corpus, the number of observations in which semantic and grammatical number do not coincide is quite considerable. Collective nouns like *hir felawshep* ‘her company’, on the other hand, are grammatically singular but semantically plural in referring to groups of several individuals. The variable number can thus be coded either as a grammatical property of the constituent or as a semantic property of the referent. Since the grammatical feature is more readily established, and since number was originally not considered as a predictor variable, the coding was ultimately restricted to the former. In terms of theoretical motivation, the assumption that the number of a constituent may impact its interpretation as more or less agentive seems plausible, however, since it is part of the hierarchy of individuation that serves as the overarching factor in Timberlake’s discussion of the replacement of genitive case for negated objects in Russian (cf. Timberlake 1977: 160–163). Singular participants are more easily conceived of as individuals, which are arguably more likely to function as agents, while less highly individuated participants would be expected to receive an agentive interpretation less easily. Apart from the difficulty of having to distinguish between semantically singular and semantically plural referents of the non-distinctive personal pronoun *ye/you*, which, in some cases, could not even be overcome by reference to the wider context of a given instance, the aspect of individuation appears to be represented more suitably by the factor abstractness, however, since the levels of this factor already distinguish between concrete and abstract referents. A variable based on number, on the other hand, would assign individual concrete arguments as well as abstract propositional arguments to the same level of singular constituents, which is probably an inferior reflection of the hierarchy of individuation invoked by Timberlake than the variable abstractness that was discussed. While number thus potentially plays a role in the interpretation of a participant as agentive, the uniform grammatical expression of semantically disparate participants prevents its immediate usefulness for the analysis.

3.4.4.3. Gender

The gender of human participants potentially plays a role in their interpretation as agents as well, since it can be surmised that, historically at least, male participants have a higher potential of being interpreted as agents than female participants (cf. the discussion in Kailuweit 2002: 76). This presumably applies to the period of English under investigation, although a claim as broad as this would probably need to be verified independently in order to avoid any circular inferences from the data. There are two more basic problems with the incorporation of gender as a variable in the present setting, however. First of all, human participants as stimulus arguments are not very frequent with certain verbs to begin with, which renders a further division into male and female participants impossible to interpret with statistical significance. As experiencer arguments, human referents are frequent, but female referents are underrepresented in both corpora, since they, with some notable exceptions, do not normally figure as characters in the literature of the *PPCME2*, and since they only gradually become more prominent among the authors of letters included in the *PCEEC*. The second basic problem is that gender, as a semantic property of the referent, is not always easy to establish, since not all nouns and pronouns automatically imply natural gender. Collective nouns, for example, can refer to groups of male or female participants, or to a mixture of both, and the referent of the anaphoric pronoun of the third person plural is, in most cases, easily identified in terms of animacy and abstractness, but not in terms of its precise gender. One could, of course, simply assume that male gender is more likely to apply in the majority of contexts of the two corpora, but such an assumption would certainly not further the purpose of detecting a differential effect of gender on the construction type. Given that the prospect of this undertaking was not particularly favourable for the reasons described, and given that gender was also not required as an agreement feature, it was not consistently coded for all verbs. The information gathered for *liken* is not entirely uninteresting, however, and allows for some general observations, which will be briefly outlined below.

First of all, the expected dominance of male experiencers in the data is confirmed. About 76.3% of all experiencers of *liken* in the dataset excluding ambiguous observations ($n = 621$) are male and only 16.1% are female. The remaining instances either involve mixed gender of coordinated participants or groups of people, or they do not exhibit natural gender, or they were not possible to classify without a significant effort that exceeded the usefulness for the present purpose. A tabular overview of the different classifications of the gender of the experiencer argument of *liken* is given in table 6 below:

gender	male	female	mixed	no gender	uncertain
frequency (%)	474 (76.3%)	100 (16.1%)	7 (1.1%)	5 (0.8%)	35 (5.7 %)

With regard to the construction type, there appears to be a certain association of female experiencers with personal constructions from E2 onwards, while male experiencers undergo a less radical shift. Any such statement can only be made with great caution, however, since it rests on a limited database and additional factors may be involved in the distribution. The frequency and proportion of female and male experiencers in impersonal and personal constructions are given in table 7 below beginning with M4, as this is the earliest period in which personal constructions involving female experiencers are documented in the data.

		M4–E1	E2	E3
female	impersonal	24 (80%)	1 (4.2%)	1 (2.8%)
	personal	6 (20%)	23 (95.8%)	35 (97.2%)
male	impersonal	228 (80.6%)	12 (15%)	0 (0%)
	personal	55 (19.4%)	68 (85%)	47 (100%)

The figures for M4 and E1 are conflated in the above table, since the low number of observations with a female experiencer in E1 affects the otherwise observable trend of the proportion of personal vs. impersonal constructions in a way that is probably not representative. Only two observations are recorded, and these are both impersonal constructions, which thus constitute 100% of all observations with female experiencers in this period. After the conflation, the proportion of personal constructions in M4 and E1 combined is quite comparable for female and male experiencers, amounting to about 20% and 19.4% respectively, while the subsequent increase in the proportion of personal constructions is more pronounced for female experiencers, with only 1 instance or 4.2% remaining impersonal compared to 12 instances or 15% of male experiencers in impersonal constructions. The final period E3 shows a complete shift of male experiencers to personal constructions, while female experiencers preserve the impersonal construction in 2.8% of all observations, although this is, again, based on a single instance and probably not representative.

The significance of the observable difference between constructions involving female vs. male experiencers in E2 remains open to question, since the low frequencies do not allow for a particularly robust statistical inference. A Fisher's Exact Test, which was employed instead of a Chi-squared test due to the fact that the expected frequency of impersonal constructions with female experiencers in E2 is below 5, does not exclude the possibility of an odds ratio equal to one with sufficient probability ($p = 0.291$). At the same time, it remains striking that virtually all instances of female experiencers in that period occur in personal constructions, while the higher proportion of male experiencers in impersonal constructions suggests that this construction remained an option even at this late stage of the development. One factor that probably contributes to this result is the fact that the observations are not entirely independent of each other but taken from letters written by a limited number of different authors, who are, in many cases, identical with the referent that the experiencer argument expresses. It is thus difficult to ascertain whether the constructions are susceptible to the experiencer's natural gender or representative of an author's inclination to use a particular construction type in their personal correspondence. The hypothesis that female authors tend to use personal constructions more consistently or earlier than male authors would be interesting to test, given the sociolinguistic design of the *PCEEC*. It might tie in with observations about the pivotal role of female language users in the spread of high-prestige changes (cf. Milroy 1980), although this is less likely for personal constructions. Alternatively, it could be related to the characterisation of the style of female authors as more involved and personal than that of male authors, which seems applicable to the feature of construction type given that the more indirect manner of expressing an experiencer's affection by impersonal construction can serve to create a conceptual distance between the speaker and the addressee. This hypothesis will not be pursued any further in the present context, however, neither in terms of experiencer gender nor in terms of author gender. The latter information is conveniently available from the *PCEEC*, which affords the tools for an inquiry into the impact of demographic features on language use. Given their relatively low frequency, impersonal verbs of emotion would probably have to be considered as part of a wider set of potential properties of gender-specific language use.

3.4.4.4. Verbal Features

3.4.4.4.1. Grammatical Properties

The central element of impersonal constructions is, of course, the verbal predicate itself. Its grammatical properties, including person, number, tense and mood, were recorded alongside

those of the experiencer and stimulus argument. Due to the morphological ambiguity of many ME and EModE verb forms, which often do not distinguish person and number outside the third person singular indicative present, this grammatical information was frequently gathered from the verbal complements rather than the verb itself, which means that control of verb agreement could not be established independently for such cases. Some ambiguity also arises from the ME spelling variants, which do not always distinguish clearly between present and past tense. In doubtful cases, the context of a given verb and the parsing information of the corpora, which includes information on non-finite forms as well as present- and past-tense forms, were consulted. Since the grammatical features of the verb were not generally considered beyond their usefulness for the establishment of verb agreement, these circumstances do not pose any fundamental problems to the investigation. It is not unthinkable that the use of a particular construction type may be influenced or correlated with certain tense forms or a particular mood of the verb, either for purely formal reasons or for functional reasons. Subjunctive forms, for example, can be used to express politeness, which constitutes a conceivable factor in the development of impersonal constructions. Although the dimension of mood may be relatable to the concept of agency, it is primarily situated on the level of speaker and addressee and not necessarily also on the level of the experiencer and stimulus argument. As such, it was not incorporated in the present study.

3.4.4.4.2. Modal Auxiliaries

An additional feature of interest in the verbal complex is the presence or absence of auxiliary verbs. This feature may be informative in the case of (pre-)modal verbs, since these provide additional information about the speaker's attitude towards the modality of a proposition. The combination of an impersonal verb with auxiliary *willen* 'to want, will', for example, can imply volition on the part of the stimulus, and, in such a case, it arguably increases the stimulus' potential for an agentive interpretation. While this would appear to be a revealing criterion for the assessment of the semantic content of an impersonal verb, there are certain obstacles to the quantified use of modal auxiliaries in a statistical model. First of all, the total number of auxiliary verbs in any of the datasets is limited. With finite active clauses of *liken*, it constitutes just under 30% of all observations or, more precisely, 219 instances or 29.4% of the set including ambiguous observations ($n = 745$) and 174 instances or 28% of the set excluding ambiguous observations ($n = 621$). Apart from passives, these figures already exclude combinations of the present participle with *ben* 'to be', which were not analysed as

grammaticalised progressive forms but as non-finite adjectival forms. In addition, combinations of the structural auxiliary *haven* ‘to have’ with the past participle, which form the present and past perfect, and those of the structural auxiliary *don* ‘to do’, including its affirmative use, need to be excluded, since they are not relevant in the present context. Their exclusion brings the total number of observations with a (pre-)modal auxiliary to 138 or about 22.2% of all unambiguous constructions of *liken*.

A closer look at these observations reveals that they are not distributed evenly among the periods. Their proportion ranges between total absence in M3 and a rather high proportion of 65.9% in E1. An overview of the absolute frequencies and the proportions of (pre-)modal auxiliaries with *liken* is given in table 8:

	M1	M3	M4	E1	E2	E3
modal	5 (14.7%)	0 (0%)	15 (7.5%)	81 (65.9%)	23 (20.5%)	14 (16.5%)
no modal	29 (84.3%)	67 (100%)	185 (92.5%)	42 (34.1%)	89 (79.5%)	71 (83.5%)

The distribution illustrates a significant gap between the early combinations of *liken* with pre-modal verbs in M1 and the later instances, which culminate in a marked predilection for the use of modal auxiliaries with *liken* in E1. The overall proportion of such auxiliaries is not interpretable as an indicator of a single semantic concept like agency, however, since it comprises a range of items relating to different types of modality. Specifically, the dataset of *liken* includes precursors of the following PDE equivalents: *can*, *could*, *may*, *might*, *shall*, *should*, *will*, *would*. Their distribution is detailed in table 9:

	M1	M3	M4	E1	E2	E3	total
can/could	0	0	0	1	4	1	6
may/might	0	0	4	64	6	0	74
shall/should	1	0	6	9	8	10	34
will/would	4	0	5	7	5	3	24

As can be seen, there is one set of items with a particularly high frequency, viz. *may* and its correlate *might*. Together, these constitute a total of 74 instances, which accounts for more than half (53.6%) of all (pre-)modal auxiliaries in the dataset and for 64 instances or 79% of all modal auxiliaries in E1. The high proportion of auxiliaries in that period is thus not a reflection of a general trend but of a specific increase of *may/might* with *liken*. The other items are less frequent overall. The second most frequent pair *shall/should* amounts to 34 instances, which corresponds to about a quarter (24.6%) of all (pre-)modal auxiliaries in the dataset and is thus much closer to the expected frequency, while *can/could* and *will/would* comprise 6 instances (4.3%) and 24 instances (17.4%) respectively. While the use of these verbs is informative for the interpretation of individual instances, a combined treatment of (pre-)modal auxiliaries for use in a statistical model of the construction of impersonal verbs would be unwarranted due to the verbs' semantic disparity. A preliminary model including the presence or absence of modal auxiliaries and the chronological period as predictors suggests a decided increase in the odds of impersonal constructions whenever the modal feature is present, and although this effect remains statistically significant if further factors such as animacy and category of the stimulus are introduced, it can be uniquely traced to the high frequency of *may/might*, which occur exclusively in impersonal constructions. For other items such as *will/would*, it seems reasonable to assume that they represent several types of modality, including volitional and epistemic modality as well as counterfactual statements, especially in view of the temporal gap between the early instances from M1 and the later instances from M4 onwards. In either case, the inclusion of (pre-)modal auxiliaries as a general factor would be dominated by the high proportion of instances of *may/might* and thus be largely representative of a single set of items. The lower frequency and disparate semantics of the other items equally speaks in favour of an individual, qualitative evaluation of this feature. While the modality expressed by modal items will thus be a possible point of discussion, they will not be considered as a quantitative variable in logistic regression.

4. Results

In the present section, the quantitative data of the impersonal verbs that were investigated will be presented and evaluated. This will be done in two consecutive steps. First, the composition of the datasets that were extracted for each verb will be described and the distribution of the annotated variable levels will be explored. Second, the influence of the independent variables on the dependent variable construction type will be investigated by fitting a logistic regression

model and by discussing any additional qualitative observations that arise from the data. This will be done separately for each verb, followed by a general discussion and an overall conclusion in the final section. Since the data were limited by a variety of factors, however, the case of ME *liken* will be the only one for which a full statistical analysis will be presented, while the discussion of the other verbs will not involve statistical inferences. These other verbs include *quemmen* ‘to please, be acceptable’, *plesen* ‘to please, satisfy’, *reuen* ‘to regret, feel pity’, *longen* ‘to desire’ and *wanten* ‘to need, lack’. As can be seen from their respective glossing, the cases that will be discussed include three near-synonymous verbs, viz. *liken*, *quemmen* and *plesen*, as well as two other verbs of emotion *reuen* and *longen*. The last item, *wanten*, is perhaps best classified as a verb of possession or lacking. Its meaning is contiguous to that of *longen*, since the lack of something can obviously imply desire, which is apparent from the central sense of PDE *want*. The comparison of different verbs facilitates an investigation of the hypothesis that the semantic roles associated with these verbs impact the interpretation of a given experiencer argument as more or less agentive, which is hypothesised to be reflected in the early adoption of nominative case marking of the experiencer argument. Verbs with identical or very similar meanings are expected to exhibit a similar development, while verbs from different semantic groups should exhibit a different development. In the present constellation, the near-synonymous verbs of liking would thus be expected to develop similarly, since their primary argument is a typical experiencer argument in all three cases. The other two verbs of emotion equally require a typical experiencer argument, however, and thus do not present maximally contrastive cases. Given the discussion in section 2.2.2.1.3, it is also not quite clear how differences in verb-based semantic roles such as repenter and desirer as opposed to liker should affect their potential in terms of an agentive interpretation. Repentance could be seen as an emotional process that requires a more active involvement of the experiencer than liking, since it presupposes a certain degree of conscious reflection, while the feeling of desire for something that is absent can, perhaps, be characterised as a more passive emotion than liking. On the other hand, both repentance and desire potentially display a higher emotional intensity than liking and thus imply a heightened degree of involvement of the experiencer in this respect. The experiencer argument of *wanten* is the only case that is more noticeably different from the other verbs, since it participates in an existential rather than an emotional experience and is therefore, perhaps, better defined as a beneficiary or an argument of pertinence (on a slightly different semantic categorisation of OE impersonal verbs cf. Möhlig-Falke (2012: 85f.), who distinguishes between existential experience for verbs like OE *gelimpan* ‘to happen’ and non-availability for verbs like OE *behofian* ‘to have need of’). The combination of verbs that were

investigated is thus not optimally suited for testing the aforementioned hypothesis that lower-level semantic roles influence the diachronic development of the impersonal verbs. Since the selection partly depends on constraints imposed by the method employed to extract tokens of the relevant verbs from the corpora, this shortcoming is difficult to address, however. Without significant pre-processing, the inclusion of other contrastive cases such as that of *thinken* ‘to think’ and its homonym *thinken* ‘to seem’, which merge into a single verb of cognition in Present-Day English, would not be possible, given the lack of adequate lemmatised corpora. The comparison of different verbs of emotion with a verb of lacking will be instructive nevertheless, even if it does not include maximally distinct cases. The order in which the verbs will be discussed is the one given above, starting with the near-synonymous verbs *liken*, *quemen* and *quemen*, proceeding with the other verbs of emotion *reuen* and *longen*, and finishing with *wanten*.

4.1. Liken

The ME verb *liken* ‘to please’ continues the indigenous verb OE *lician*. It is a verb of emotion that is typically constructed as a type-ii verb in Old English and Early Modern English, but also participial constructions are attested from Old English onwards (cf. the entries in Bosworth/Toller 1898: 637 and 1921: 616). The diachronic development from an oblique-experiencer verb to a nominative-experiencer verb makes the assumption of two separate verbal meanings, viz. causative ‘please’ and receptive ‘like’ (cf. Fischer/van der Leek 1983: 341), a convenient descriptive tool. Nevertheless, the theoretical motivation of multiple verbal meanings and the question of whether the verb’s diachronic development actually implies a semantic shift or whether a single general meaning should be assumed is debated. In the following section, the dataset that was extracted for *liken* will first of all be described in terms of its composition of finite, non-finite and passive clause types, the presence of explicit experiencer and stimulus arguments, and the degree of ambiguity of the construction type. In the subsequent section, the quantitative and chronological distribution of the semantic and syntactic properties that function as potential predictor variables will be outlined.

4.1.1. Database

While many of the figures relating to the dataset of *liken* were already cited in the relevant sections on the subsetting of the data and the annotation of the variable levels, it will be

convenient to reiterate these figures here for systematic reference and comparison with the other verbs. The total number of observations cited below already excludes instances that were extracted from the corpora but either not identifiable as tokens of *likem* with certainty or obvious false positives. While the number probably gives a more or less accurate approximation of the actual frequency of this verb in the two corpora, the possibility remains that certain graphemic variants were not detected by the script that was used to extract the forms. This is already apparent from the fact that at least one additional instance of *likem* was discovered by pure chance during the annotation process. The verb's frequency in the two corpora is, of course, not representative of all varieties and strata of the English language by equal measure, although it constitutes the best estimate that can be achieved by the present means. References to the frequency of *likem* thus always need to be understood with these two caveats in mind.

The full dataset of *likem* contains 831 observations, whose distribution across the seven ME and EModE periods that were investigated is given in figure 2:

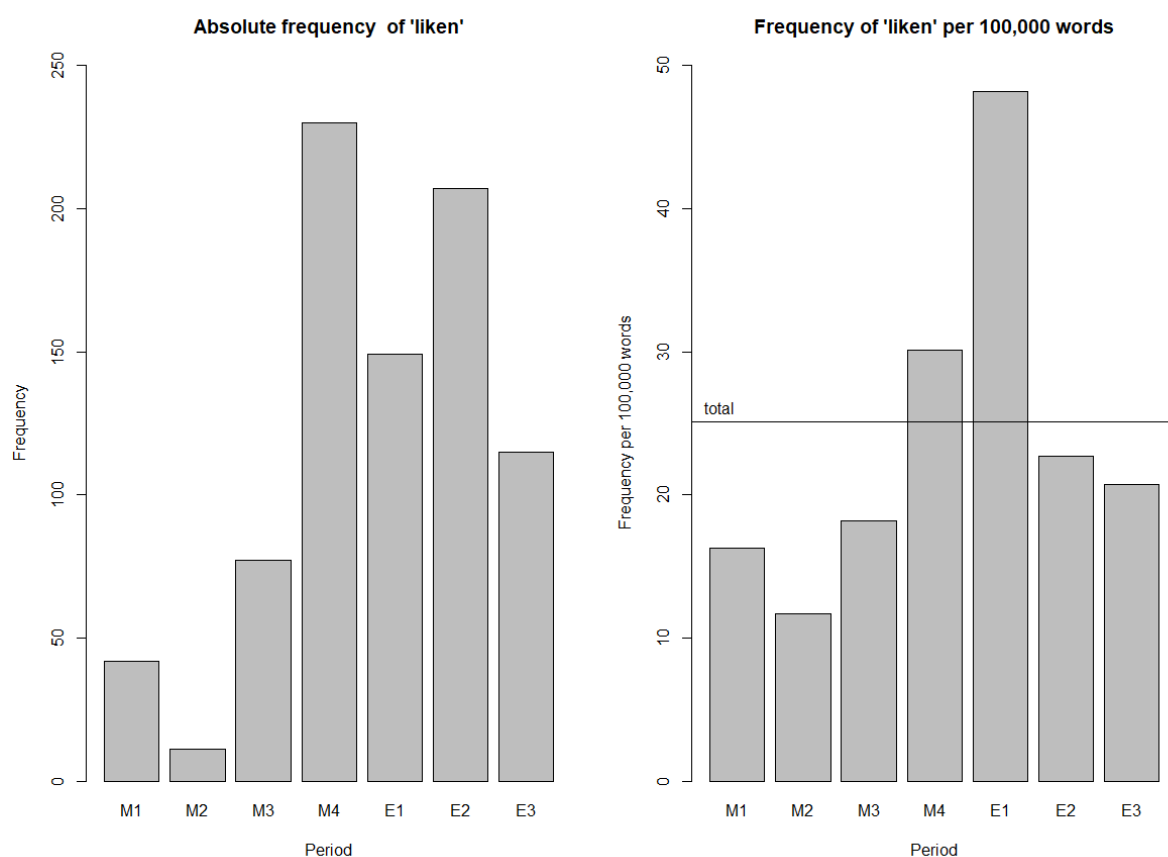


Figure 2 Absolute and standardised frequency of *likem* by period (n = 831).

As can be seen from the plot on the left-hand side, most of the data of *likem* derive from the later periods M4–E3. Together, these constitute 701 or about 84.4% of all 831 observations, while the earlier periods M1–M3 only amount to 130 or about 15.6% of the total data. This

uneven distribution of the verb's frequency is, of course, partly related to the fact that the corpora contain an uneven amount of data for each period, which is particularly sparse in M2 and particularly rich in E2 (cf. table 1 in section 3.1 for a detailed breakdown of the word frequencies). In order to obtain an idea of the diachronic frequency of *likēn* as a lexeme, its frequency was standardised to the number of occurrences per 100,000 words. The resulting values are displayed in the plot on the right-hand side and indicate that *likēn* is infrequent also in standardised terms in M2, which contains only 11.7 occurrences per 100,000 words, while the highest standardised frequency is observed in E1 with 48.2 occurrences per 100,000 words. Both of these values differ markedly from the standardised total frequency of *likēn* in the corpus, which is represented by the horizontal line in the plot. Incidentally, it should be noted that the frequency of *likēn* in M2 is based on its occurrence in three text samples only. In all of these, its frequency is below par, but this need not be representative of the period in general, for which more extensive documentation is absent. There is also a certain unevenness in the distribution of *likēn* across different text samples. Since some of the collections of letters in the *PCEEC* span across several periods, a text-based comparison of the frequencies from a single period is not always possible when data from that corpus are involved. It is still noticeable that several of the collections in E1 exhibit a particularly high frequency of *likēn*. For example, in *The Correspondence of Sir Thomas More*, this frequency amounts to 194.9 occurrences per 100,000 words and in *The Letters of William, Lord Paget of Beaudesert* it amounts to 102.2 occurrences per 100,000 words. The relatively high frequency of *likēn* in E1 is thus not characteristic of all texts in that period, but the fact that several collections contain a much higher number of instances per 100,000 words than the corpus as a whole does indicate that at least some the authors of that period display a decided propensity towards the use of *likēn*.

If the level of individual text samples is considered more closely, it becomes apparent that 422 or more than half (51.8%) of all occurrences of *likēn* derive from only 10 different samples, while the remaining 409 occurrences (49.2%) are distributed across 91 text samples. The samples with the highest proportion derive almost exclusively from the *PCEEC*, which is mainly due to the large size of some of the collections in that corpus: the *Paston Letters and Papers* alone contain 234,098 words and are thus almost four times as large as the largest sample in the *PPCME2*, which is drawn from *The Book of Margery Kempe* and contains 60,212 words. In addition, some of these text samples display a particularly high standardised frequency of *likēn*. This includes the correspondence of Thomas More and William Paston mentioned above, but also the *Stonor Letters and Papers* and the *Letters of Dorothy Osborne to William Temple*, which contain an extrapolated number of 128.9 and 71.7 instances per

100,000 words respectively. Text samples with a high incidence of *likēn* in the *PPCME2* include *Julian of Norwich's Revelations of Divine Love*, the *Cloud of Unknowing* and Chaucer's *The Tale of Melibee* with frequencies of 279.9, 89.7 and 76.4 instances per 100,000 words, but even though the first of these frequencies is exceptionally high, the small sample size of only 5,004 words means that its impact on the total number of observations remains relatively limited. Using Gries' standardised Deviation of Proportions as a measure for the dispersion of *likēn* across the two corpora (cf. Gries 2008, Lijffijt/Gries 2012), it emerges that the extreme preference of certain text samples for the use of *likēn* contributes noticeably to the distribution of the verb, which is moderately uneven according to the value obtained from the comparison of expected and observed proportions within the text samples ($DP_{\text{stand}} = 0.37$, where 0 represents a maximally even distribution according to the sample sizes and 1 a maximally uneven distribution, in which all occurrences are contained in the smallest text sample). Of course, the largest text samples continue to have the greatest impact on the database, unless their standardised frequency of the verb is particularly low. Since the larger collections of the *PCEEC* often contain letters written by several authors across a period of several decades, the impact of such large samples does not seem to constitute a fundamental problem, however.

As discussed in section 3.3 of the methodology, the quantitative analysis of verbal tokens was restricted to finite active clauses, since non-finite instances and passive clauses may not exhibit the same complementation patterns. The proportion of the different clause types and their distribution across the seven periods is displayed in figure 3:

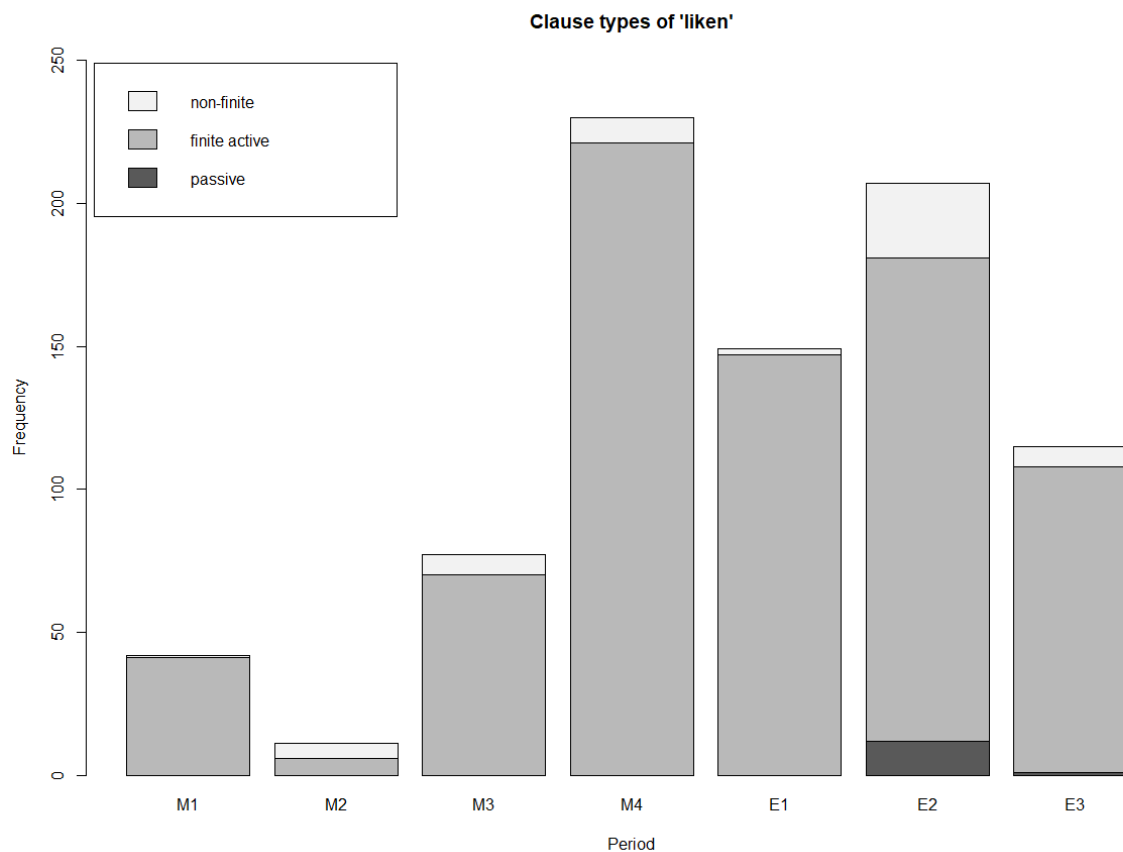


Figure 3 Clause type of *liken* by period (n = 831).

Taken as a whole, finite active clauses constitute 761 or 91.6% of all 831 observations of *liken*, while non-finite instances and passive clauses constitute 57 observations (6.9%) and 13 observations (1.6%) respectively. Regarding their diachronic distribution, finite active clauses constitute the clear majority in all periods except M2, where their majority is less pronounced. Only 6 or 54.5% of the 11 observations in that period involve finite active clauses, while the remaining 5 observations (45.5%) are all non-finite. Given the low number of observations in M2, this distribution may, of course, be coincidental. In all other periods, the proportions of non-finite instances are more moderate, ranging between 1.3% in E1 and 12.6% in E2. The comparatively high incidence of non-finite forms of *liken* in E2 is not easily traced to a single factor, since non-finite forms occur in a variety of contexts, which include subject-to-object raising, participial clauses and infinitival complements of nouns and adjectives, all of which are attested in this period. Passive clauses do not occur in the data until E2, where they constitute 12 or 5.8% of 207 observations, and their number drops to only 1 additional occurrence in the subsequent period E3. With 81.6%, E2 thus presents the period with the lowest proportion of finite active clauses after M2, which is mainly due to an increase in the number of alternative constructions, while finite active clauses remain a stable clause type with *liken* throughout the investigated period. The subset of 761 finite active clauses was further reduced by the exclusion

of observations that did not contain both arguments of the verb. For *likēn*, this involves 8 observations with a missing experiencer argument and another 8 observations with a missing stimulus argument, leaving a total of 745 observations.

The final reduction of the dataset concerns observations that were ambiguous in terms of their construction type or in their expression of one of the predictor variables. The latter case was only found with a single observation of *likēn*, which contained two coordinated stimulus arguments that did not coincide in their level of animacy. The remaining 744 observations are displayed in figure 4, which illustrates the proportion of ambiguous and unambiguous observations in each period.

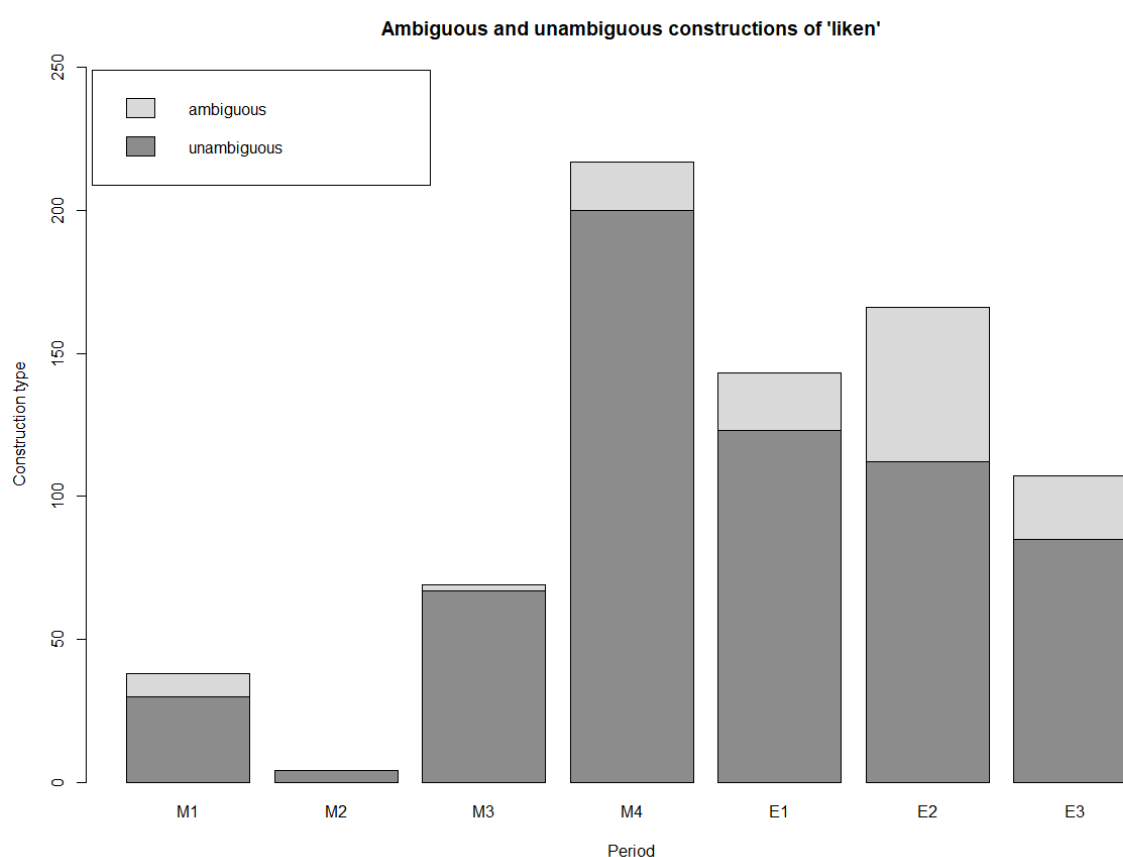


Figure 4 Ambiguity of construction of *likēn* (n = 744).

As can be seen, the only period without ambiguous observations is M2, and this complete absence may be coincidental, since only 4 observations remain in that period after the exclusion of non-finite instances and observations with missing arguments. When looking at the proportion in the earliest periods M1 and M2 combined, ambiguous observations constitute 8 or about 19.0% of all 42 observations, which is very similar to the proportions in the latest period E3, where 22 or about 20.6% of all 107 observations are ambiguous. There is thus no obvious diachronic trend in morphosyntactic ambiguity. Nevertheless, the lowest proportion is

found in the ME period M3 with only 2 ambiguous observations (2.9%) of a total of 69 observations, while the highest proportion of ambiguous constructions is observed in E2, where 54 observations (32.5%) of a total of 166 observations are ambiguous with regard to their construction type. The overall proportion of ambiguous observations across all periods amounts to 16.5%, which may seem surprisingly high given the fact that experiencer arguments are often assumed to be realised by morphologically unambiguous pronouns. While this is certainly true to a large extent, case marking is not a pervasive factor in the present data, which becomes even more apparent when taking the fact into account that only 477 or about 64.1% of the 744 observations are disambiguated by case marking or the equivalent use of a PP experiencer, while 144 or 19.4% are disambiguated by one of the other criteria of construction type, especially by constituent order in *it*-constructions. Two factors contribute to the high proportion of ambiguous constructions in the later periods, viz. the loss of distinctive case marking in the second-person plural pronoun *you*, which is particularly influential in the correspondence corpus, and the use of full NPs as abstract terms of address, which is also fairly common in the parts of the correspondence corpus. Although the latter instances are frequently disambiguated by the use of an expletive, the exclusion of ambiguous observations significantly impacts the size of the database. Taking all of the cases discussed above into account, the reduction is particularly noticeable in the data for E2, which contains 207 observations in the initial dataset and is reduced to 112 observations after the removal of non-finite instances, passives, instances with missing arguments and ambiguous constructions. The following sections will outline the diachronic distribution of the construction type and that of the predictor variables.

4.1.2. Variables

4.1.2.1. Construction Type

The final dataset of unambiguous instances of *likēn* contains 621 observations, 370 (59.6%) of which are constructed impersonally according to one of the criteria described in section 3.4.1 above, while the remaining 251 (40.4%) are constructed personally. The distribution of the construction types across the ME and EModE periods is displayed in the two plots in figure 5, where the right-hand plot shows the proportion of the construction types in each period and the left-hand plot shows the absolute frequencies on which these proportions are based. The importance of absolute frequencies for the interpretation of proportions is immediately apparent from the situation in M2, which contains only 1 personal and 3 impersonal constructions. This would suggest a proportion of personal constructions of 25%, which is much higher than in the

two adjacent periods. Since this figure cannot be regarded as accurate in view of the low number of total observations, the data for M1 and M2 are conflated in the proportions plot in order to obtain a slightly more representative picture:

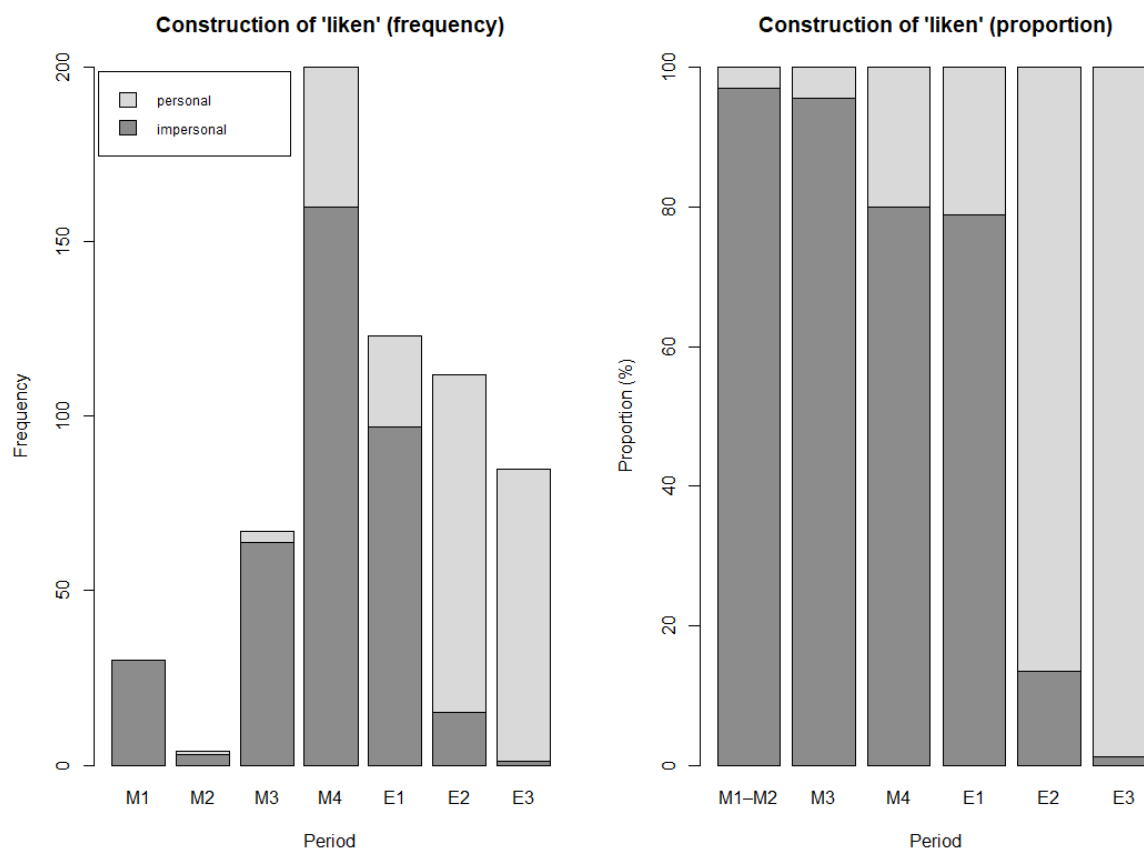


Figure 5 Frequency and proportion of the construction of *liken* (n = 621).

As can be seen in the right-hand plot, the development of the construction of *liken* shows three distinct diachronic stages. During the first stage (M1–M3), impersonal constructions are the dominant construction type with only 1 personal construction (2.9%) in M1–M2 and 3 personal constructions (4.5%) in M3. The second stage exhibits a moderate increase in the proportion of personal constructions to about 20.0% in M4, which is sustained in E1 at about 21.1%. During the final stage (E2–E3), the initial situation is reversed, and personal constructions constitute the clear majority of all constructions with 86.6% in E2 and 98.8% in E3, in the latter of which only a single impersonal construction remains.

The most notable fact about the diachronic development of the construction type is its relative stagnation during the second stage (M4–E1). This situation is in stark contrast to the expected S-curve, in which an exponential increase of the innovative construction during the first half would be followed by a logarithmic increase during the second half of the development. There is a specific factor that can explain the sustained proportion of impersonal

constructions, however, and that is the rise of (mostly) impersonal constructions with an expletive constituent. If these are excluded from the data, this not only reduces the absolute number of observations in the central periods M4 and E1 significantly, it also affects the proportion of the construction types in those periods in a way that the result resembles an S-shaped development much more closely. Figure 6 displays the distribution of construction types without *it*-constructions. Specifically, observations including expletive (*h*)*it* or, once, *that* were removed, more than half of which also contain an unambiguously case-marked experiencer argument. Stimulus arguments realised by propositions, which are by far the most frequent type of complement in these constructions, also occur outside of *it*-constructions, but even though certain formulaic expressions containing propositional arguments were classified as impersonal constructions based on their similarity to *it*-constructions, these were not removed from the dataset without *it*-constructions on which the plots are based.

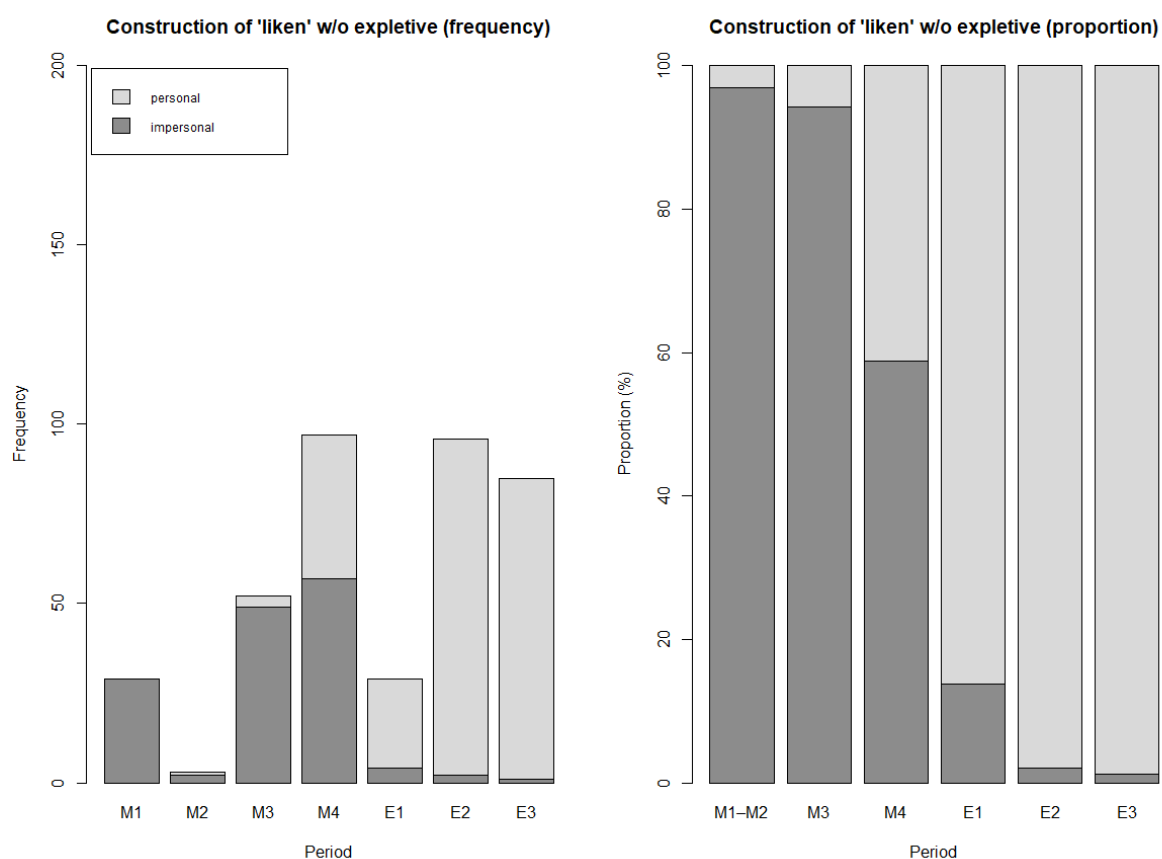


Figure 6 Construction of *liken* without expletive (n = 391).

The left-hand plot shows the reduction of observations in absolute terms, which is most noticeable in M4 and E1, where 103 observations (51.5%) and 94 observations (76.4%) of the total number of observations contain an expletive. Since these constructions are almost exclusively impersonal, the remaining cases exhibit a significantly larger and more rapid

increase in the proportion of personal constructions, which constitute around 41.2% of all observations in M4 and 86.2% in E1. The diachronic development from impersonal to personal constructions is thus much more similar to an S-curve when considered separately from constructions containing an expletive. The distribution and construction type of observations that include an expletive are plotted in figure 7 for comparison.

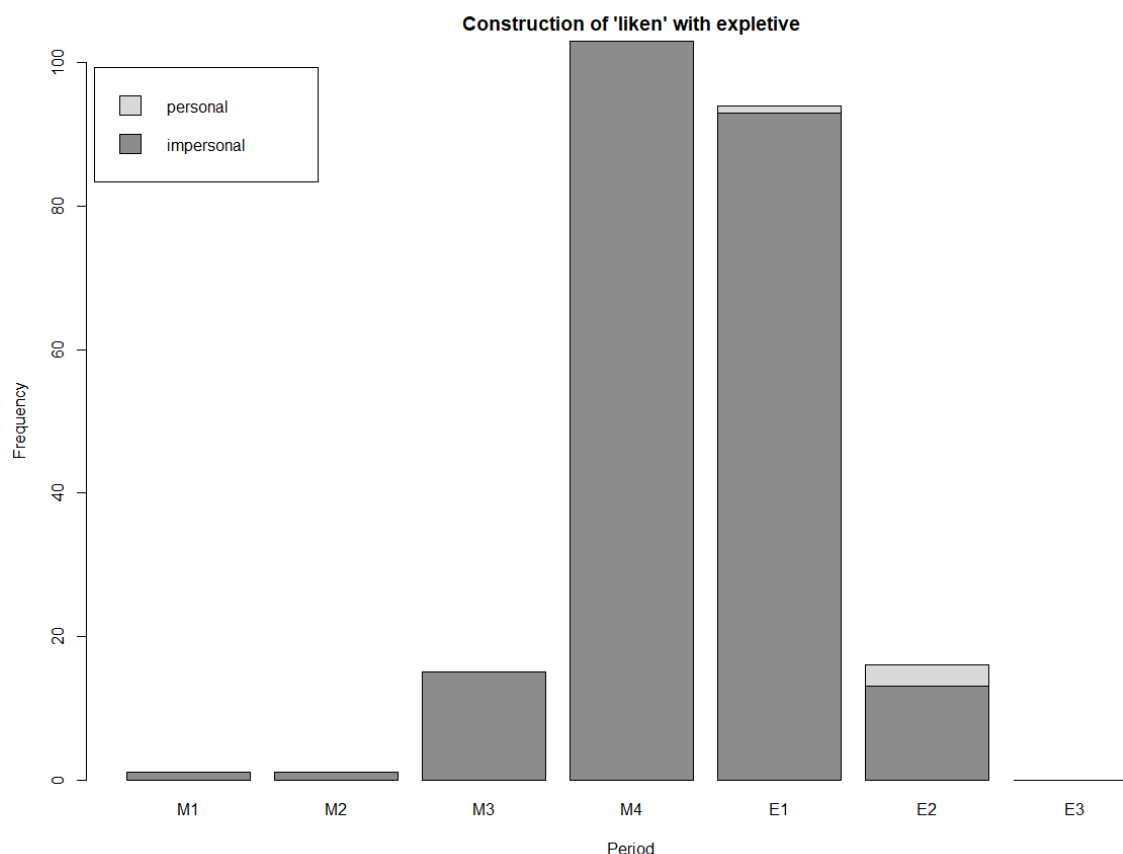


Figure 7 Construction of *likēn* with expletive (n = 230).

Since the constructions with an expletive are also almost exclusively impersonal, the proportions of the two construction types are not plotted separately. The light-grey areas in E1 and E2 of the frequency plot indicate, nevertheless, that a small number of *it*-constructions were identified as personal, since the pronominal constituent in these cases occupied a position comparable to that of other objects (cf. e.g. *I like it not amis that you haue stayed his younger brother [...]* (PASTONK,81.051.973)). The function of the pronoun can either be identified as that of an empty expletive or as referring cataphorically to the subsequent object clause, but since the same interpretation could be applied to instances in which the pronoun fills the position of a subject, this does not discount the pronominal objects from inclusion. Even though the personal constructions comprise only 1 instance in E1 and 3 in E2, they illustrate that the presence of an expletive is not necessarily equivalent to an impersonal construction. The plot

also shows that the use of expletives is concentrated around the central periods M4 and E1, while M1 and M2 contain only a single instance each and E3 does not contain any instance of an expletive pronoun. Overall, the emergence of constructions with an expletive and their predominantly impersonal construction type adequately account for the sustained number of impersonal constructions observed in the data for M4 and E1.

4.1.2.2. Predictor Variables

4.1.2.2.1. Animacy

The animacy of the stimulus argument is one of the two central semantic variables whose influence on the construction type the present investigation is intended to test. As discussed in section 3.4.2.1, two levels of animacy were distinguished for this purpose, viz. non-human and human stimulus arguments. A closer look at the data reveals that these two levels are distributed very unevenly for *liken*, both in general quantitative terms and in diachronic terms. Among the 621 observations, only 55 stimulus arguments (8.9%) are human, while the remaining 566 (91.1%) are non-human. The distribution is displayed below in figure 8:

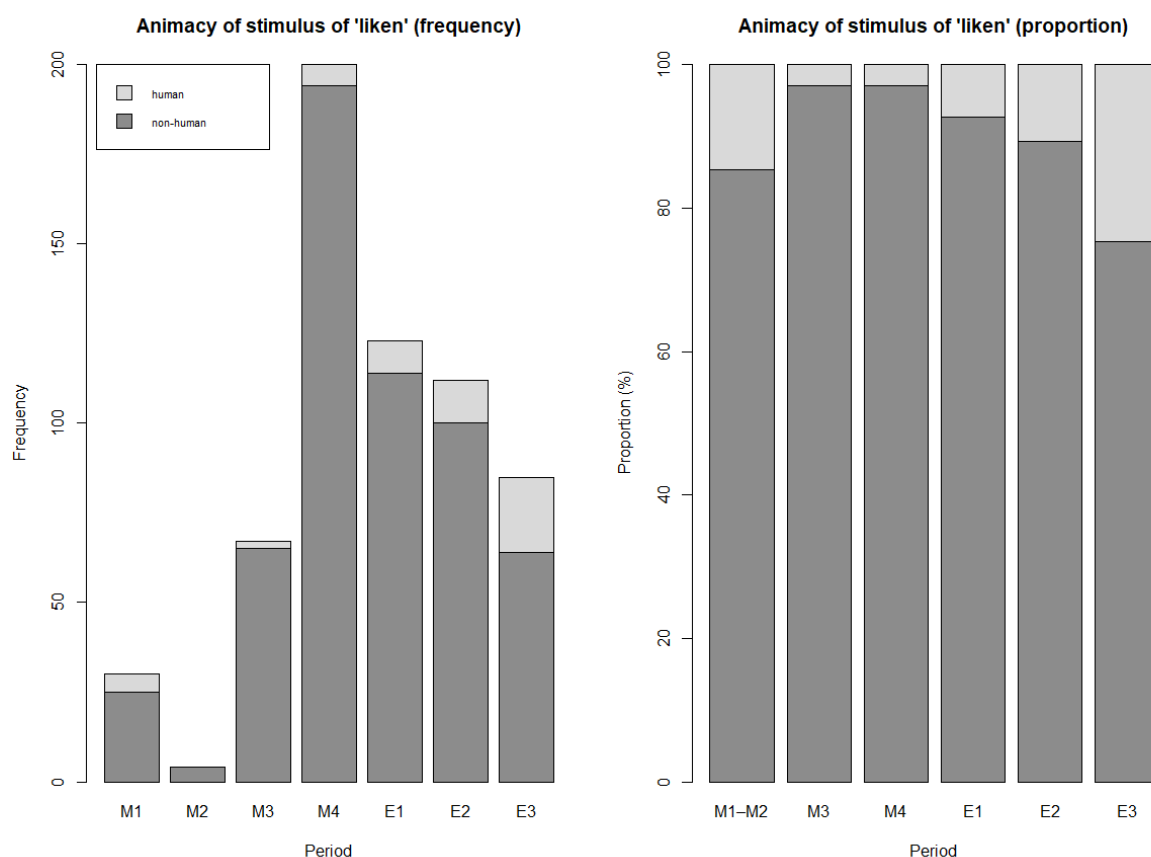


Figure 8 Animacy of the stimulus argument of *liken* (n = 621).

With the exception of 5 early instances in M1, the proportion of human stimulus arguments shows a clear diachronic trend. It starts out low at the level of 3.0% in M3 and M4 and continually increases at a moderate rate to 7.3%, 10.7% and 24.7% in periods E1, E2 and E3 respectively. Also the absolute frequencies of human stimuli are low, ranging between 2 instances in M3 and 21 instances in E3. While the quantitative basis for assessing the influence of animacy on the construction type thus becomes more robust in the later periods, these are also the periods in which personal constructions are largely dominant and the dependent variable shows less variation. Any observed association of human stimuli with personal constructions could thus be due to an impact of animacy on the construction type or to the late emergence of human stimuli with *liken* at a time when personal constructions were already the norm, which makes it difficult to distinguish between the elapse time and the semantic property of animacy as the most relevant factor. The predominant use of non-human stimuli with *liken* during the earlier periods in particular is an interesting property, however, which contrasts both with the PDE situation and the situation of *quemen* in Middle English.

4.1.2.2.2. Abstractness

The second semantic property of the stimulus argument that will be presented is abstractness. For this variable, three levels were distinguished, viz. concrete stimuli, abstract stimuli and propositional stimuli (cf. section 3.4.2.2). As with the partly correlated levels of animacy, the quantitative distribution of these categories is rather uneven. Propositional stimuli constitute the absolute majority with 371 instances (59.7%) of all observations, while abstract and concrete stimuli constitute 112 instances (18.0%) and 138 instances (22.2%) respectively. The preponderance of semantically propositional stimuli is not surprising, since they occur in virtually all observations including an expletive, but they also occur in a range of other observations without an expletive. Concrete and non-propositional abstract stimuli, on the other hand, although less frequent, are certainly not as marginal as human stimuli. The diachronic distribution of the three levels of abstractness is displayed in figure 9:

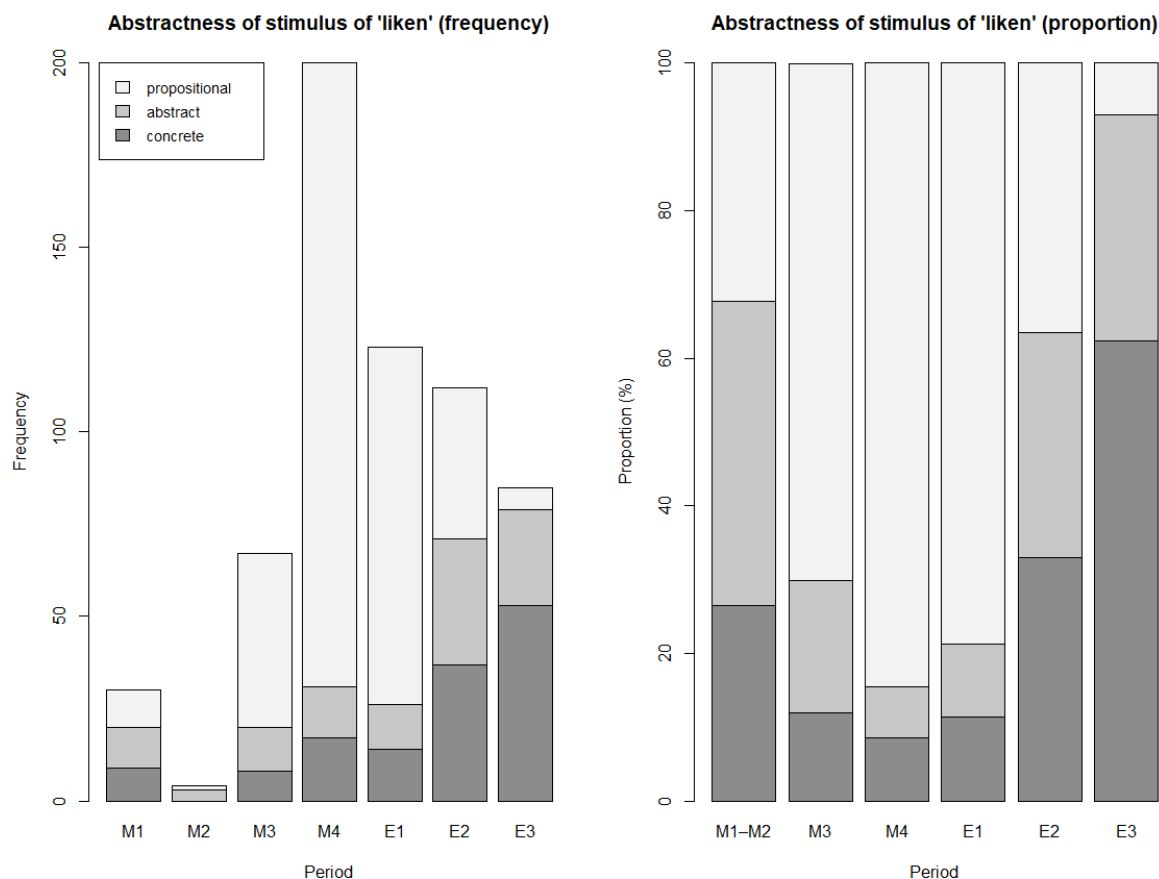


Figure 9 Abstractness of the stimulus argument of *likēn* (n = 621).

As can be seen from the right-hand plot, the high proportion of propositional stimuli is particularly pronounced during the middle periods M3, M4 and E1. This is in keeping with the high number of constructions with an expletive in the latter two periods (cf. figure 7 above), but these cannot account for the high proportion of propositional stimuli in M3, which exhibits much fewer observations with an expletive. They also do not account for all observations in M4, where 68 (40.2%) of 168 observations with a propositional stimulus in terms of abstractness do not contain an expletive. Propositional stimuli are thus not exclusively tied to the increase of *it*-constructions of *likēn*. The proportion of the other two levels of abstractness reflects the high frequency of propositional stimuli during the middle periods and is thus highest during the early periods M1 and M2 and the late periods E2 and E3. When compared against each other, an additional increase of concrete stimuli compared to abstract stimuli is also noticeable. The former constitute 17 (39.5%) of all 43 non-propositional stimuli in M1–M3 and 53 (67.1%) of all 79 non-propositional stimuli in E3. There is thus an observable trend towards more concrete stimuli in the data, which also constitute the absolute majority (62.4%) of all 85 observations in the latest period E3.

4.1.2.2.3. Category

In terms of lexical and syntactic realisation, six different categories were distinguished for stimulus arguments, viz. full NPs (i.e. NPs with nominal heads), personal pronouns, other pronouns, PPs, deleted constituents (“null”) and propositions (cf. section 3.4.2.3). It should be remembered that personal pronouns were defined as pronouns of the first and second person and of the third person singular masculine and feminine, and that propositions include both finite and non-finite clauses, either as explicit constituents or as constituents inferred from the context. The distribution of these categories, which is plotted in figure 10, illustrates, once more, the impact of constructions with a propositional stimulus argument on the total database:

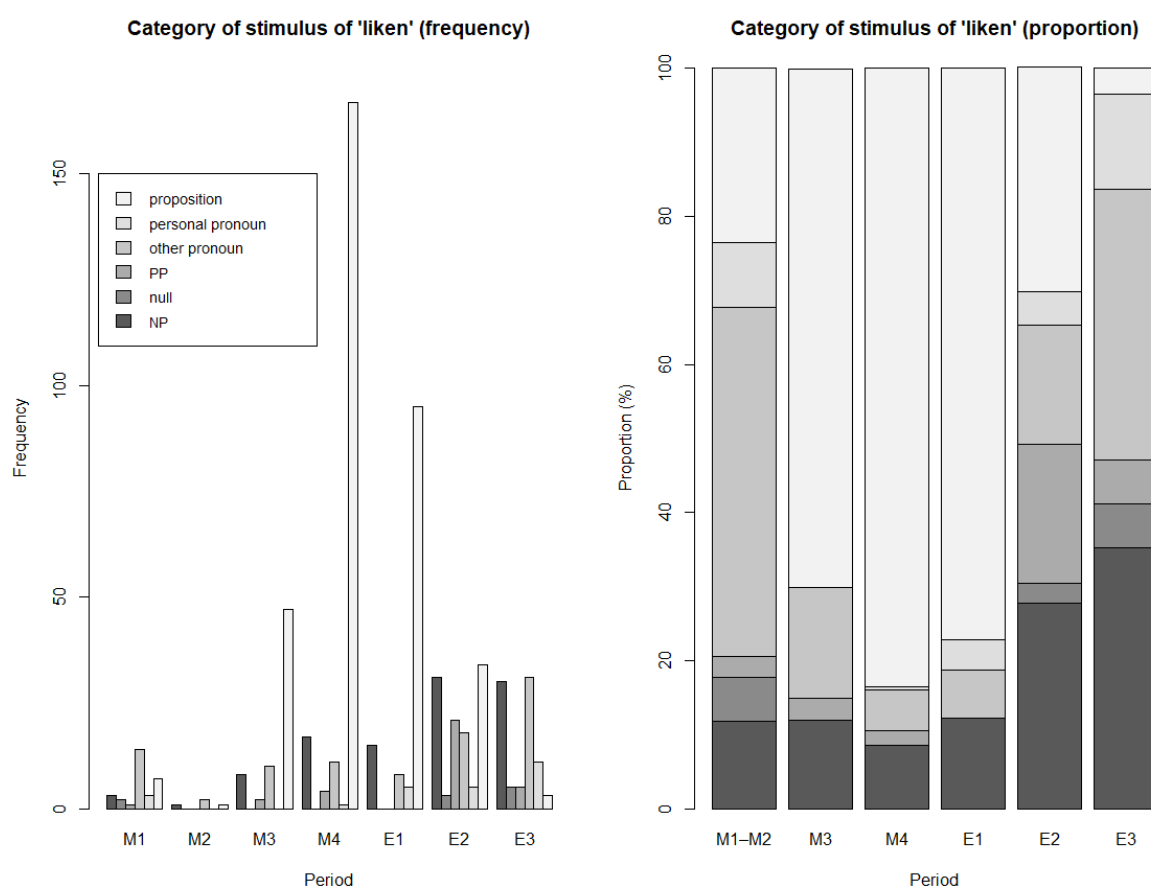


Figure 10 Category of the stimulus argument of *liken* (n = 621).

The proportions of stimuli that are propositional in terms of category are, of course, almost identical to those of propositional stimuli in terms of abstractness (cf. figure 9 above), since the latter involve only a few additional observations that result from cases in which an anaphoric pronoun refers to a propositional antecedent. The dominance of propositional stimuli during the middle periods M3–E1 is thus once more observable in figure 10. The proportion of syntactically propositional stimuli is only somewhat lower in M1–M2 and E2, where 3 instances

(27.3%) and 7 instances (17.1%) of propositional stimuli in terms of abstractness are not propositional in terms of category. Before turning to the other categories, a few remarks on propositional stimuli will be instructive, since these can be viewed from a number of different angles, including their conjunction with an expletive and their status as inferred constituents.

As was already noted, constructions with an expletive constitute a large share of the observations with propositional stimuli in most periods. For comparison, the frequencies of propositional stimuli without an expletive are displayed in figure 11.

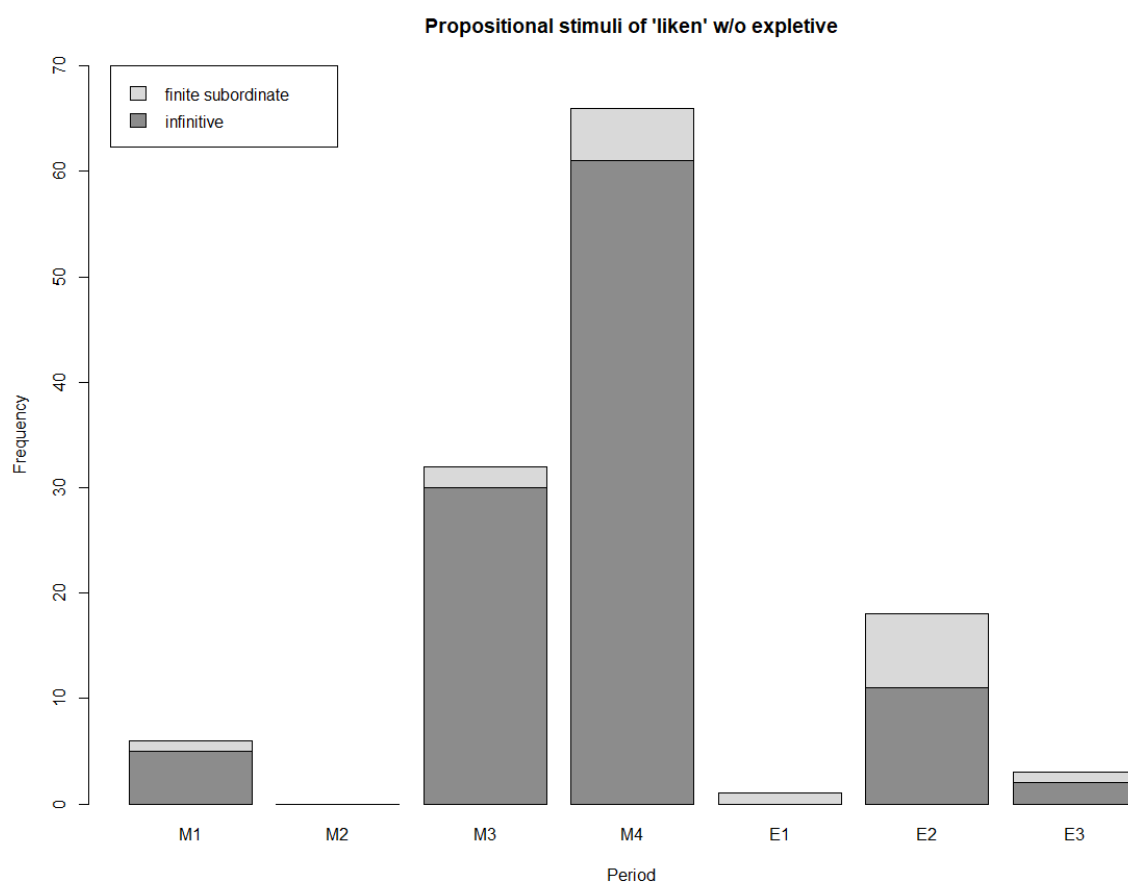


Figure 11 Propositional stimulus arguments of *likēn* without expletive (n = 126).

Since the absolute number of observations with a propositional stimulus argument is very low in M1–M2 and E3, the contrasting proportions of propositional arguments with and without an expletive are not plotted separately. A comparison with the plot of the total frequencies of propositional stimuli in figure 10 reveals, however, that the middle periods M3–E1 differ greatly in this respect. While 32 (68.1%) of 47 observations with propositional stimuli in terms of category do not contain an expletive in M3, their proportion falls to 66 observations (39.5%) out of 167 in M4 and to just 1 observation (1.1%) out of 95 in E1. There is thus a significant number of observations with propositional stimuli outside of *it*-constructions, but their proportion declines notably during the three periods described. The plot in figure 11 also

illustrates that the majority of propositional stimuli are infinitival clauses. This is true for propositional stimuli regardless of the presence of an expletive, even though the frequencies of infinitival complements are only plotted here for observations without an expletive. Infinitival clauses also constitute the majority of instances in which the stimulus argument is not explicit but inferred from the context. In absolute terms, such inferred stimuli are the most frequent in M3 and M4, where they constitute 28 (59.6%) and 31 (18.6%) of all 47 and 167 propositional stimuli respectively. This makes them the most typical type of propositional stimulus argument in M3, although their proportion is even higher in M1, where all 7 instances of propositional stimuli are inferred, and certain instances also occur in E2 and E3. Within M3, the most common clause type that contains inferred stimulus arguments are comparative clauses (cf. e.g. *[a]nd he hath as many paramours as him lyketh [sc. to have]* (CMMANDEV,24.582)), which account for 16 (57.1%) of the instances in that period, and which belong to the clause type referred to by Allen as “variable clauses” (cf. Allen 1995: 339 and section 2.2.1.2.1).

The diachronic distribution of the five remaining categories is not so easy to describe in broad terms, since their frequency per period is generally much lower than that of propositional stimuli. The least frequent non-propositional category are deleted constituents, which comprise 10 (1.6%) of all 621 observations, as well as personal pronouns, which comprise 25 (4.0%) of all observations. Since personal pronouns were defined as pronouns with exophoric reference to typically animate referents, their low frequency is obviously related to the low frequency of human stimuli, which exhibit a moderate increase towards the end of the investigated period. The proportion of personal pronouns increases accordingly, amounting to 11 instances in E3, which corresponds to 13.4% of all non-propositional stimuli in that period. The next most frequent category are PPs. Their highest proportion is observed in E2, where they constitute 21 instances or about 26.9% of all 78 non-propositional stimuli, while most other periods exhibit only sporadic instances of PPs as stimulus arguments. The use of PPs as stimulus arguments thus appears to be more of a temporary fashion than a consistent option with *liken*. Finally, the most typical realisation of the stimulus argument of *liken* besides clausal constituents are full NPs and non-personal pronouns. Together, these comprise between 62.8% and 90% of all non-propositional items across the periods, and although their diachronic distribution is not entirely even, there are enough representatives of these categories in each period to allow for an investigation of their potential influence on the construction type by logistic regression. The following section 4.1.3 will give an outline of how the influence of the variables described in the preceding paragraphs was represented by logistic regression.

4.1.3. Logistic Regression Model

The semantic and syntactic variables for which the observations were annotated were tested as predictors in a logistic regression model of the odds of impersonal and personal constructions. Such a model allows for the representation of the association of multiple independent predictor variables with a dependent variable. The odds of the response variable are estimated as log odds or “logits”, which are centred around zero. Positive estimates indicate that the odds of the compared response level are greater than the odds of the response’s reference level, while negative estimates indicate that the odds of the compared level are lower than the odds of the reference level in a given constellation of predictor variables. In the present case, this means that positive estimates indicate greater odds of personal constructions and negative estimates indicate greater odds of impersonal constructions, since the latter were chosen as the reference level of the construction type. The estimates of the predictor variables are given in log odds ratios, which are also centred around zero. Positive estimates indicate an increase in the odds of the compared response level, while negative estimates signify a decrease in these odds.

Due to the categorical nature of the response variable construction type, the present case of logistic regression requires the verification of fewer assumptions than linear regression, but there still are a couple of issues that need to be addressed. First, the observations of a dataset should normally be independent of each other in order to ensure that variation in the response variable is indeed relatable to variation in the designated predictor variables and not to interindividual variation between the subjects from which the observations are drawn. In practice, this criterion is often relaxed in linguistic corpus studies, however, since it is often difficult to avoid drawing more than one observation from any given author or text in a corpus (cf. Levshina 2015: 271). The problem becomes less severe in larger samples with more varied data, but it remains a potential factor when working with historical data, whose availability is limited, and which cannot necessarily be sampled in a way that avoids interdependence of the observations. In the full dataset of *liken*, for example, as many as 103 observations were derived from a single collection of letters, and although the collections in the *PCEEC*, for the most part, contain letters by a variety of different authors, certain authors take a greater prominence than others, which may have an influence on the overall results. One way to address this is by including clustering factors such as the identity of the author or text as random effect in a mixed model in addition to the fixed effects that a regression model is intended to estimate, but this approach was not taken in the present study (cf. section 4.1.3.1.4 below).

The second issue to be considered is that of collinearity between the predictors. Its presence can lead to unreliable estimates, since the effect of a given predictor cannot be estimated independently if it covaries with another predictor. Collinearity is particularly relevant in the present case because the semantic and syntactic variables for which the data were coded represent concepts that partially overlap. Participants that are human in terms of animacy will always be concrete in terms of abstractness and never abstract or propositional. On the other hand, there is a perfect association of arguments that are propositional in terms of category, i.e. realised by finite or non-finite clauses, and arguments that are non-human or propositional in terms of animacy and abstractness. Clearly, the inclusion of all three predictors in a single model would lead to different degrees of redundancy. This can either be addressed by changing the configuration of the variable levels in a way that reduces the overlap of the predictors, or by selecting a combination of the available variables that avoids the duplication of covarying features of the participants as much as possible. Since the overarching function of the observable predictors is to represent a given participant's potential for an agentive interpretation and to assess the influence of such a potential on the diachronic development of the construction type, theoretical considerations are also important for the selection of the variables and their configuration in terms of level distinctions. The configuration of the available predictors and their combination in the regression model of the construction of *liken* will be discussed in more detail in section 4.1.3. The general presentation and discussion of the results obtained for the impersonal verbs under discussion will be the focus of the following section.

4.1.3.1. Configuration

4.1.3.1.1. Periods

As discussed in section 3.4.3, the period information associated with each observation was included in the regression model in order to control for the aspect of the diachronic change in the odds of the two construction types under investigation. By default, this implementation of time as a categorical variable takes the form of a comparison of each factor level to a specified reference level, which is most naturally set to the earliest period in the data. Since the order of periods is not arbitrary but sequential, it would also be possible to specify an ordered factor rather than an unordered factor. Such a specification would not impact the estimates of the theoretically important variables, however. The advantage of an unordered factor, on the other hand, is that its estimates are readily interpretable and easier to evaluate than the linear,

quadratic and cubic etc. estimates that would be obtained for an ordered factor. The maximum detail of an unordered factor consists in the distinction of all seven periods represented in the corpora, viz. the four ME periods M1–M4 and the three EModE periods E1–E3. For the data of *likem*, a minimal conflation of M1 and M2 was required, however, since the absence of personal observations in M1 constitutes a quasi-complete separation of the control variable and the dependent variable. This creates a problem for the regression model, since the response can be perfectly predicted from the reference level, to which the other predictor levels can no longer be meaningfully compared. The diachronic control variable at its maximum level of detail thus takes the form of an unordered factor with six periods and the conflated periods M1–M2 as its reference level.

The extremely uneven diachronic distribution of the levels of the response variable or, in other words, the fact that the process under investigation takes the form of a near-complete shift from impersonal to personal constructions implies that associations between the predictors and the response variable are established by the regression model based on a poor representation of one of the response levels in the marginal, i.e. the earliest and latest, periods. The greater detail of a factor including the maximum number of period divisions is therefore offset by a potential lack of statistical significance when comparing periods that differ only moderately from the reference level, as well as by relatively large standard errors and correspondingly large confidence intervals, which affect all period comparisons and the intercept if the reference period is kept short. These even affect the comparison of the final period E3, although this period will be significantly different from the reference level given that it presents a reversal of the initial distribution of the construction types. While a precise estimation of the coefficients of the control variable is not of primary concern for the study, the additional advantage of further period conflations in order to ensure a more robust representation of the response levels in each period lies in a simplification of the model at the cost of only marginally increased residual deviance while preserving relative stability of the coefficients for the theoretically important variables. Of course, this can only be achieved as long as period conflations do not cancel out the effect of vastly different periods.

A first step in assessing the rate by which the odds of the two construction types change over time is the manual inspection of their diachronic proportions, which are plotted in figure 5 in section 4.1.1.2.1 above. Based on this plot, the change from impersonal to personal constructions of *likem* can be roughly divided into three phases, viz. an early phase (M1–M3) with mostly impersonal constructions, a middle phase (M4–E1) with impersonal and personal constructions at a ratio of about 4 to 1, and a late phase (E1–E2) with mostly personal

constructions. This manual division of the periods can be checked against the estimation of the coefficients for each period in a regression model that contains the maximum number of period levels as the sole predictor. The model shows that the estimated coefficients for M4 and E1 are very similar to each other (2.1102 and 2.1799) but markedly different from those of the earlier period M3 (0.4362) and the later periods E2 and E3 (5.3632 and 7.9273). Taking the logarithmic nature of the coefficients into account, where increases between lower values correspond to a higher increase in the corresponding odds than increases between higher values, this essentially confirms the relative homogeneity of the phases that result from the manual division. A second option is to assess the period divisions in terms of their ability to discriminate significantly between resulting sets of observations and their respective association with the response variable. This can be done with the help of a conditional inference tree on the basis of the same model, using the `ctree()` function of the *party* package in *R*. The algorithm of this function recursively partitions the independent variables according to their association with the response variable, starting with the strongest association, if any, and ending with the last association that meets a number of criteria including a specified significance level, which is set to $p = 0.05$ by default. The partitions resulting from the application of this algorithm coincide with the manual divisions, but in addition to the division of M1–M3 and M4–E1, the split between E2 and E3 is found to be significant ($p = 0.002$). Given that the standard error and confidence interval of the coefficient for E3 remain relatively large, and given that the improvement of the overall model is limited while the coefficients for the conceptual variables remain relatively constant, the simpler model was preferred over the one that contains an additional division between E2 and E3. This model thus includes three distinct periods, an early period M1–M3, a middle period M4–E1, and a late period E2–E3, which constitutes the baseline to which the conceptual variables were added.

4.1.3.1.2. Conceptual Variables

Besides the diachronic control variable, three conceptual variables were coded for the purpose of logistic regression. Since the experiencer argument is relatively constant in its attributes, they all relate to the stimulus argument, specifically its animacy, abstractness, and the lexical or syntactic category by which the argument is realised. It is obvious that these variables constitute representations of partially overlapping features, which is why the selection of variables for the regression model and the configuration of their constituting levels need to take potential redundancies and the option of multiple perspectives into account. In principle, the distinction

of variable levels can proceed from two fundamentally different starting points. On the one hand, the distinctions should be based on theoretical considerations, which present the motivation of the hypothesised influence of a given variable. On the other hand, an observed similarity in the effect of given variable levels may motivate the conflation of these levels, even if this was not preconceived theoretically. A reduction of the number of variable levels can result in improved statistical significance of the remaining distinctions, since these will be represented by a higher number of data points, as well as a simplification of the overall model. This data-driven approach obviously needs to be balanced by the interpretability of the resulting theoretical distinctions, and, in an ideal case, will yield significant effects of theoretically motivated distinctions only.

4.1.3.1.2.1. Levels of Category

The above considerations are particularly relevant to the configuration of the variable that encodes the lexical and syntactic categories of the stimulus. On the one hand, this is the most complex predictor in the regression model, comprising six different categories, some of which already constitute conflations of more fine-grained category distinctions. At the same time, the data for *likēn* that remain in the subset of unambiguous constructions in finite active clauses do not represent all of these categories equally well. More importantly, though, the associations of the category levels with different degrees of the overarching construct of agentivity is not immediately apparent for all categories. As discussed in section 3.4.2.3.2, the distinction of personal and non-personal pronouns, for example, was based on their referential properties, which, in the case of personal pronouns, coincides with the property of animacy of the referent. On the other end of the spectrum, propositional arguments, which were defined as finite and non-finite clauses in either overt or inferred form, coincide with the most abstract level distinguished for abstractness of the stimulus. In both cases, the grammatical realisation constitutes a reflection of a semantic property of the referent, but the mapping of categories to particular semantic properties is less direct in the remaining cases of NPs, PPs, non-personal pronouns and deleted constituents. A possible step in simplifying the variable would be to conflate the two pronominal categories with deleted constituents, since these have a comparable anaphoric function, but whether this would result in an improvement of categorial realisation as a predictor of the construction type used with impersonal verbs is not immediately apparent from a conceptual point of view.

In order to test categorial realisation as a variable in the context of the limitations of the dataset of *likem*, a simple regression model including this variable at its maximum level of detail and the period distinctions as control variable was fitted using the `glm()` function in *R*. By the default setting, all levels of a factor are compared to a single reference level, whose specification thus influences the type of comparisons made as well as the resulting coefficients and significance levels. Given the theoretical uncertainty surrounding the association of the category levels with different degrees of agentivity, the definition of a reference level to which the other categories should be compared is not immediately obvious. A particularly strong association with the response variable or poor quantitative representation by a low number of observations are not advantageous properties of a reference level, however, since, in the former case, all resulting comparisons are likely to emerge as significant based on the strength of a single association, and, in the latter case, none of the resulting comparisons may emerge as significant, since the other levels cannot be confidently compared to a poorly represented reference level. Different settings of the reference level were hence applied to the variable of categorial realisation, and these unanimously identified the contrast between propositional arguments and all other categories as the only contrast at the significance level of $p < 0.05$. The only exception to this were deleted constituents, which are represented by only 10 observations, and for which no significant contrast with any other category, including propositions, could be established.

While none of the other categorial distinctions revealed a contrast at the desired significance level, a profile of the associations of the different category levels with the response variable can nevertheless be obtained from the multiple comparisons with different reference levels that were made. Deleted constituents are disregarded, since they are not only few in number but also distributed across the periods in a way that all instances fall either into the early or the late phase, in which their association with either impersonal or personal constructions is less informative. The remaining categories fall into the following cline of increasingly strong associations with personal constructions displayed in (151):

(151) NPs < non-personal pronouns < PPs < personal pronouns

The ranking of the categories is based on the sum of their coefficients in the specified regression model with different settings of the reference level. For example, the log odds ratios of personal and impersonal constructions for NPs compared to non-personal pronouns, PPs and personal pronouns are estimated at -0.1486, -1.5164 and -1.6519 respectively. The sum of these estimates is -3.3169, which is the lowest value for any of the categories. The equivalent sums of the

coefficients obtained for the three remaining categories are -2.7225, 2.7487 and 3.2907. It should be pointed out, once more, that these values are only indicative of the relative association of the compared categories with the response variable, which means that the negative values obtained for NPs and non-personal pronouns do not necessarily signify a strong association with impersonal constructions. In the present case, these values merely indicate that NPs and non-personal pronouns are less strongly associated with personal constructions than PPs and personal pronouns, but all four categories are strongly associated with personal constructions when compared to propositional arguments in the dataset. The second point to bear in mind is that moderate differences in association require larger amounts of data in order to be statistically significant. As is to be expected, the data available for *likem* are insufficient for validating the relatively small differences between NPs and non-personal pronouns and between PPs and personal pronouns, whose absolute z-scores and associated p-values of 0.295 ($p = 0.768$) and 0.1355 ($p = 0.919$) fall well below the level of significance, and even the most pronounced difference between NPs and personal pronouns is only marginally significant ($p = 0.100$).

The question remains whether more robust statements about the impact of categorial realisation can be made based on theoretically meaningful confluences of the level distinctions that cannot be verified individually. The most obvious step based on the associations observed in the data would be to conflate the levels with similar associations, viz. NPs and non-personal pronouns on the one hand and PPs and personal pronouns on the other hand. While the former conflation could, perhaps, be justified in terms of the non-restrictedness of NPs and non-personal pronouns in their reference to animate or inanimate participants, which was also used as a criterion in the distinction of personal and non-personal pronouns (cf. section 3.4.2.3), the latter conflation seems difficult to motivate from a theoretical point of view. Instead, a conflation of PPs with NPs and non-personal pronouns would seem to be more plausible. Their stronger association with personal constructions would remain unexplained, however, and the overall effect of the resulting category would become less pronounced. The same is true for the theoretically plausible conflation of the two pronominal categories (and deleted constituents), whose differing associations with the construction type would be neutralised in the resulting category of pronouns, which, again, would make the difference in its association compared to other categories more difficult to substantiate. Indeed, no significant contrasts emerge from any of the aforementioned category confluences, although this is probably at least partly due to limited quantitative representation in the case of personal pronouns, which represent the category that is most strongly associated with personal constructions in the data. Overall, only one categorial distinction is thus to be carried over from the present discussion, viz. that

between non-propositional and propositional stimuli. The latter of these display a strong relative association with impersonal constructions compared to the other individual categories as well as all other categories subsumed as non-propositional stimuli as a whole.

4.1.3.1.2.2. Levels of Abstractness

The second predictor variable for which multiple levels were coded is the degree of abstractness of the stimulus, which includes concrete, abstract, and propositional stimuli. The configuration of this variable is more straightforward than that of categorial realisation, since its levels fall into a natural cline from more concrete to more abstract participants, and since all levels are relatively well represented in the data of *liken*. Conflation for the sake of an improved quantitative representation of a variable level against which the other variable levels can be compared is thus not a necessary issue here. Both of the possible conflations, which result in a distinction of concrete vs. abstract or of non-propositional vs. propositional stimuli, would be easy to motivate from a theoretical point of view, although the attendant reduction in the level of detail of the variable may not be desirable for the purpose of interpretation. Given that the levels of abstractness partly overlap with the levels of animacy and category, the main question, however, is which configuration of the levels of abstractness can improve a model that already contains the other variables. Since concrete stimuli include all human stimuli, and semantically propositional stimuli are virtually coextensive with the syntactic category of propositions defined in section 3.4.2.3.2, the choice available for this end is very limited. Even though semantic propositions include a small number of additional cases of propositions realised by anaphoric pronouns and, in two instances, by PPs, the inclusion of both terms in the same model would create unnecessary redundancy. This means that a selection among the available variables is required in order to adequately model the diachronic change from impersonal to personal constructions.

4.1.3.1.2.3. Selection of Variables

While the number of available regression terms is not particularly large to begin with after the conflation of non-propositional stimuli as a single level of the variable category, the preceding discussion suggests that further simplifications are required in order to address the redundancy of the predictors. The procedure applied to this end is similar to that of the reduction of category levels and involves a preliminary regression model in which different settings of the reference

level ensure that all potentially relevant comparisons are made. The conceptual variable entered into the model besides the diachronic control variable constitutes a conflation of the semantic variables animacy and abstractness. Given that human stimuli are a subset of concrete stimuli, animacy and abstractness can be easily merged into a single variable that represents both dimensions. With this configuration, all comparisons between the four resulting groups of human, non-human concrete, abstract, and propositional stimuli are immediately accessible and thus more easily interpretable than the coefficients obtained for a combination of multiple dummy variables with partially overlapping levels. This presents a methodological advantage, even though the overlap between two dummy variables would also be accounted for by the regression model. For example, the coefficient established for human stimuli compared to non-human stimuli in a model containing animacy and a conflated version of the levels of abstractness with non-propositional and propositional stimuli as predictors is identical to that for human stimuli compared directly to non-human stimuli in a single variable that distinguishes human, non-human (i.e. concrete and non-propositional abstract) and propositional stimuli. Taking the semantic variable with four distinct levels as a starting point, the sums of the coefficients resulting from the comparison of all possible level contrasts were calculated. These suggest a cline of association with personal constructions of the kind given in (152):

(152) human > concrete > abstract > propositional

The cline is in agreement with the theoretically expected structure of the variable, since human participants are considered to be the most concrete level and propositional arguments the most abstract level of the variable. The rounded sums of coefficients are 6.27 for human stimuli, 3.26 for (other) concrete stimuli, 0.12 for (non-propositional) abstract stimuli and -9.65 for propositional stimuli, which indicates that all four groups have relatively distinct associations with the response variable construction type.

As with the levels of category discussed earlier, these sums of coefficients should not be taken to automatically imply that the observed differences between the groups are statistically significant and allow for inferences beyond the present dataset. The individual level comparisons do indicate, however, that all of the differences are attested well enough in the dataset to warrant further consideration. Similar to the syntactic variable of categorial realisation, the most pronounced difference between semantically propositional stimuli and all other categories is also the most significant one with $p > 0.001$ for all possible comparisons, which is an understandable finding given the strength of the association and the high quantitative representation of propositional stimuli in the data. The differences between the

remaining groups need to be considered in two separate steps. Human and abstract stimuli differ significantly ($p = 0.020$) when compared as individual groups. This is understandable, since the difference between them is also the most pronounced among the three remaining groups. The intermediate group of non-human concrete stimuli does not differ significantly from either human or abstract stimuli, however, when compared individually against these groups ($p = 0.3012$ and $p = 0.1598$ respectively). The distinction of concrete stimuli does become significant, however, when these are contrasted with abstract stimuli under the inclusion of the subgroup of human stimuli. Similarly, human stimuli can be meaningfully contrasted with non-human stimuli if both non-human concrete and non-human abstract stimuli are included in the latter level, although the significance of this contrast is somewhat weaker than the comparison between human and abstract stimuli as individual groups. The main conclusion of this discussion is that both the dimension of animacy and the dimension of abstractness are eligible candidates for a model of the diachronic development of impersonal constructions, but the data are not strong enough to facilitate a fine-grained distinction that also includes the intersection of these two dimensions. Since animacy and abstractness are both significant when considered in isolation, the only way to represent both variables at the desired significance level is thus by two separate regression models.

The final point to consider is the inclusion of regression terms involving the syntactic variable of categorial realisation. Since the groups of semantically and syntactically propositional stimuli are virtually coextensive, this is a question of substitution rather than addition. The partial collinearity of the variables including these terms can be quantified in the form of their variance inflation factors (VIF), which can be obtained from a model containing both terms side by side using the appropriate function from the *car* package in *R*. The rounded scores for abstractness and category returned by this function are 8.52 and 8.32 respectively, and although different rules of thumb exist as to what values constitute a cause for concern (cf. Levshina 2015: 272), their exceedance of a moderately conservative threshold of above 5 sufficiently illustrates the partial overlap between the two variables. A similar correspondence exists between personal pronouns and human stimuli, but the overlap between the variable levels, in this case, is not mutual. All personal pronouns refer to human participants, a restriction that follows from their very definition, but not all human participants are realised by personal pronouns. Given that the observable association of personal pronouns with the personal construction of *liken* thus rests on a much smaller quantitative basis than that of human stimuli, it is not surprising that the contrast between personal pronouns and the other lexical and syntactic categories, apart from propositional arguments, was already found to be non-

significant in section 4.1.3.1.2.1 and hence does not need to be considered further as a candidate for the substitution of a semantic regression term.

The other level of syntactic realisation that needs to be considered are propositional stimuli, which are almost coextensive with semantically propositional stimuli. A comparison of different versions of a regression model that contain propositional stimuli either in terms of their categorial realisation or in terms of the abstractness of their referent shows that the former generally perform better in terms of reducing the residual deviance (about 1.22–1.24% of the null deviance of 837.94). However, the concordance index *C*, which indicates the proportion of instances in which the model predicts a higher probability for the construction type actually encountered, is raised only marginally by 0.002–0.003 points. While the improvement is thus not entirely negligible, the performance of the semantic regression term is not much worse and close enough to consider the option of maintaining a model based solely on semantic predictors. The question is whether the categorial realisation of stimulus arguments as propositions presents a variable that is independent of the semantic features of the referent and, if so, how the given association of this particular category can be accounted for. Alternatively, the categorial realisation could be regarded as an expression of the underlying semantic feature of the referent, which would then constitute the more appropriate explanatory variable from a theoretical point of view. The apparent difference between the predictive power of a model containing syntactic propositions as a regression term compared to a model containing semantic propositions as a regression term would need to be explained, however. Ultimately, the semantic version of the variable of propositional stimuli was chosen as the primary candidate for the regression model, since no other level of the variable category exhibited a significantly strong association with the response variable, and since the syntactic category of propositional stimuli constitutes the level of categorial realisation most closely correlated with the semantic properties of the referent. Besides the distinction of human and non-human stimuli (or of concrete and non-concrete stimuli), the regression model thus contained the distinction between semantically propositional stimuli and semantically non-propositional stimuli as the second conceptual regression term.

4.1.3.1.3. Interactions

The next step in the configuration of the regression model was to consider whether any of the variables displayed significant interactions that would need to be integrated. In general terms, this would take into account whether the effect of a given predictor variable on the odds of the

response levels changes depending on the level of another predictor variable. In the present scenario, this is only relevant for potential interactions of the conceptual variables with the diachronic control variable, since the former represent distinct levels of a single variable for which interactions cannot possibly be observed. The effect of human vs. non-human stimuli on the odds of the construction types, for example, cannot differ depending on whether the stimulus is also propositional or non-propositional, since both human and non-human stimuli are non-propositional by their definition as the intersection of non-human and non-propositional stimuli. In other words, they do not constitute independent variables but dummy variables representing the levels of a single variable. A difference in the effect of these variables depending on the period of the diachronic control variable is still conceivable, however, since the associations of the represented levels with the construction type need not be constant over time. To some degree, this is even expected as part of the design of the regression model, in which the diachronic dimension of the data is represented in the form of discrete phases, since all constructions in the dataset of *likēn* are initially impersonal and thus display identical associations of the semantic levels with the construction type, which remain overwhelmingly impersonal and therefore largely indistinct in the selected reference period. The differences only become more pronounced in the central phase and generally converge again towards the end of the investigated period, as personal constructions become the dominant construction type for all semantic types of the stimulus argument. An integration of these facts can thus enhance the precision of the regression model, which potentially leads to additional insights about the effect of the semantic variable.

A methodological problem with the implementation of interactions in small datasets is that they require at least one observation of each response level for every combination of the levels of the interacting variables, which is difficult to ensure given the limited amount of available data. For example, human and concrete stimuli are not attested in personal constructions in the dataset until M4, which results in a quasi-complete separation of the response levels in the earliest phase M1–M3, since human (and concrete) stimuli perfectly predict the construction type in those periods. In order to be able to meaningfully compare the difference in association of human and non-human stimuli to that in other periods, the period needs to be expanded so as to include at least one personal and one impersonal observation for each variable level. Based on the diachronic distribution of human stimuli in the present dataset, a minimum conflation of M1–M4 on the one hand and of E2–E3 on the other hand is required, with the additional necessity of conflating E1 with either one of these periods. It is obvious that such confluations result in a significant loss of precision and, more importantly, in a reduction

of diachronic control, since the diachronic distribution of the different types of stimuli within each larger period is no longer taken into account. While a comparison of the effects in periods involving large-scale conflation is thus less accurate and even potentially spurious, the degree of accuracy in the final period E2–E3 is identical to that achieved by the general model. The isolated effect for that period can be obtained either by implementation of the interaction or by subsetting the data to this period. The observed effect can then be compared to the effect of the variable estimated across all periods. At the same time, the final period E2–E3 obviously constitutes the period of convergence of both construction types into personal constructions, in which less pronounced and, consequently, less robustly attested differences between observations with different types of stimuli are generally expected. Their interpretation thus rests on an integration with additional facts from the complete dataset, which is precisely the advantage of the full regression model.

Focussing on the difference between human stimuli and non-human stimuli in E2–E3, it emerges that the association of non-human stimuli with personal constructions is actually stronger in that period than that of human stimuli, which is the reverse of the average effect observed across all periods. The relationship is visualised in the two plots in figure 12 below, which were created using the `visreg()` function from the *visreg* package in *R*:

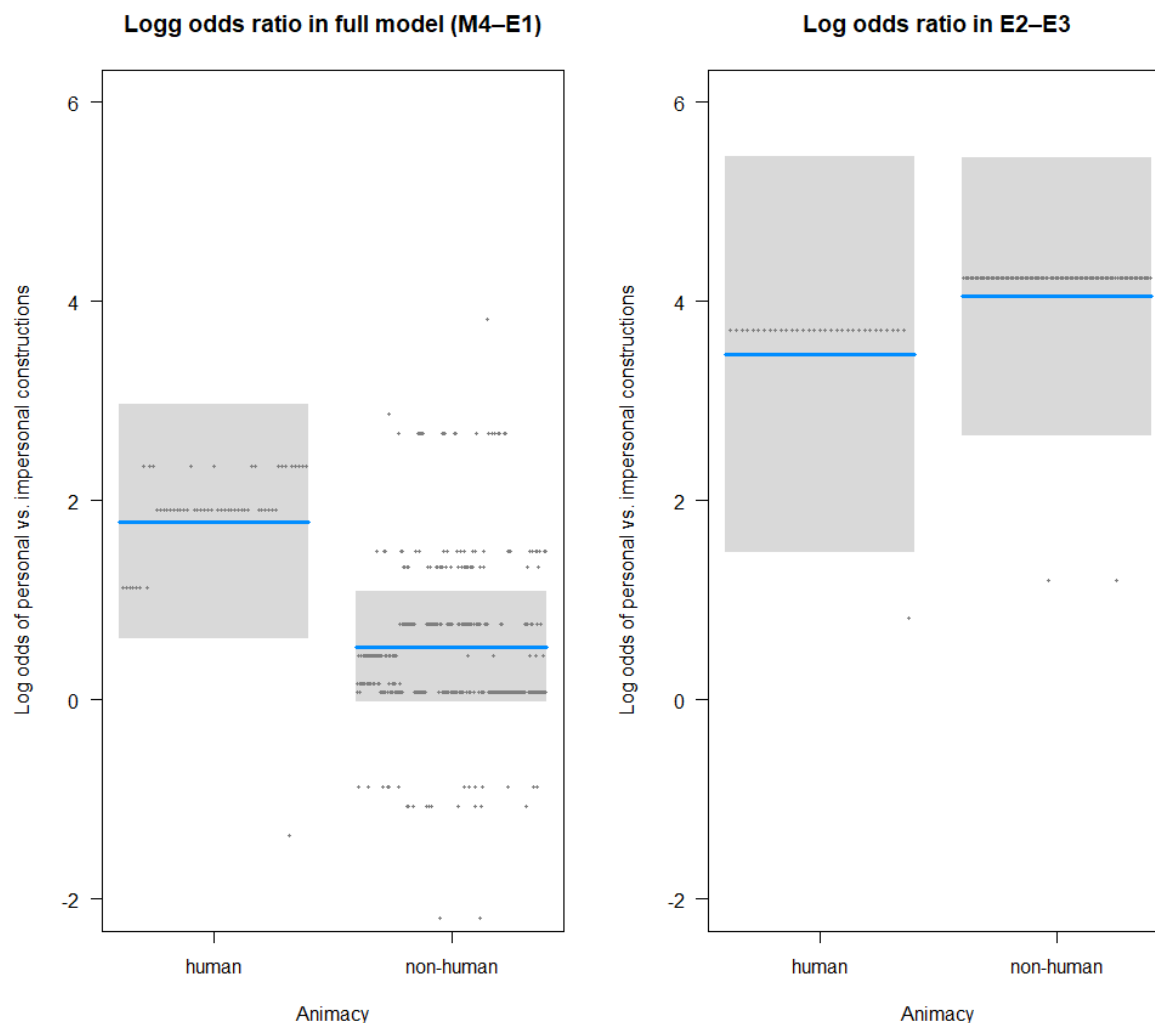


Figure 12 Contrast of human vs. non-human stimuli of *liken* in full model vs. E2–E3.

The plot on the right-hand side visualises the log odds ratio of personal vs. impersonal constructions for human compared to non-human stimuli in E2–E3. It is based on a simple regression model that excludes propositional stimuli as well as all observations from the earlier periods M1–E1 ($n = 150$). The left-hand plot visualises the log odds ratio obtained from the full regression model including the ternary period division and propositional stimuli, although the log odds of the latter are not visualised. The specific log odds presented are the ones that obtain in the central period M4–E1. As can be seen, the log odds ratios differ in sign. Non-human stimuli increase the log odds of personal constructions in E2–E3 by a coefficient of 0.586 but decrease the log odds in the full model by an estimated coefficient of -1.2548. In contrast to the effect in the full model, the effect in E2–E3 is not significant ($p = 0.64$), which can also be inferred from the large confidence bands rendered as grey boxes in the plot. It is nevertheless necessary to ask whether the impact of animacy on the construction type is a valid effect if it appears to be reversed over time. The frequencies, which, incidentally, can be read from the

number of partial residuals represented as dots in the same plot, give a good indication of why the association of non-human stimuli with personal constructions is deemed to be stronger than that of human stimuli. While both semantic groups exhibit almost no instances of impersonal constructions in the final period E2–E3 (2 and 1 instance respectively), non-human stimuli are attested in personal constructions a lot more frequently than human stimuli (115 instances compared to 32 instances). It is, of course, not easy to tell whether a greater sample of human stimuli would result in a similar distribution or not. A more detailed look at the proportion of personal constructions in the semantic subgroups shows that (non-human) concrete stimuli are most strongly associated with personal constructions in the period under discussion (100% personal), while the proportion is somewhat more moderate for abstract stimuli (96.7% personal) and human stimuli (97.0% personal). The log odds ratio resulting from this constellation for (human and other) concrete stimuli compared to abstract stimuli (-1.121) is almost identical to that in the full model (-1.0996) and does not need to be visualised separately. Overall, then, the complete absence of impersonal constructions involving (non-human) concrete stimuli and the relatively high frequency of observations involving personal constructions account for the apparent reversal of the association of human vs. non-human stimuli in the final period E2–E3.

The development of the effect of propositional stimuli vs. non-propositional stimuli across the different periods can be tested somewhat more straightforwardly, since the former are relatively well documented in both construction types for most periods. The only conflation necessary are those of the reference period M1–M3 and of the final period E2–E3. The reference level is presented by non-propositional stimuli, which can be defined as either non-human (i.e. concrete and abstract) or non-concrete (i.e. abstract) stimuli, and these are attested at least once in personal and impersonal constructions for each of the resulting periods in either version. In order to circumvent the unstable coefficient obtained for human vs. non-human stimuli in an interaction model with more than two period levels, which, as discussed earlier, is a result of the separation of the response levels by human stimuli in at least one of the periods, the interaction of the contrast of propositional and non-propositional stimuli can be visualised independently on the basis of a subset excluding human stimuli or of a model with two separate dummy variables, to each of which the appropriate interactions can be added. The identical coefficients resulting from either approach are visualised in the plot in figure 13 below. Since a more fine-grained period distinction yields additional information, a display of the interaction using a quaternary period distinction is given, which makes the best possible use of the available data. The contrasted periods are M1–M3, M4, E1 and E2–E3:

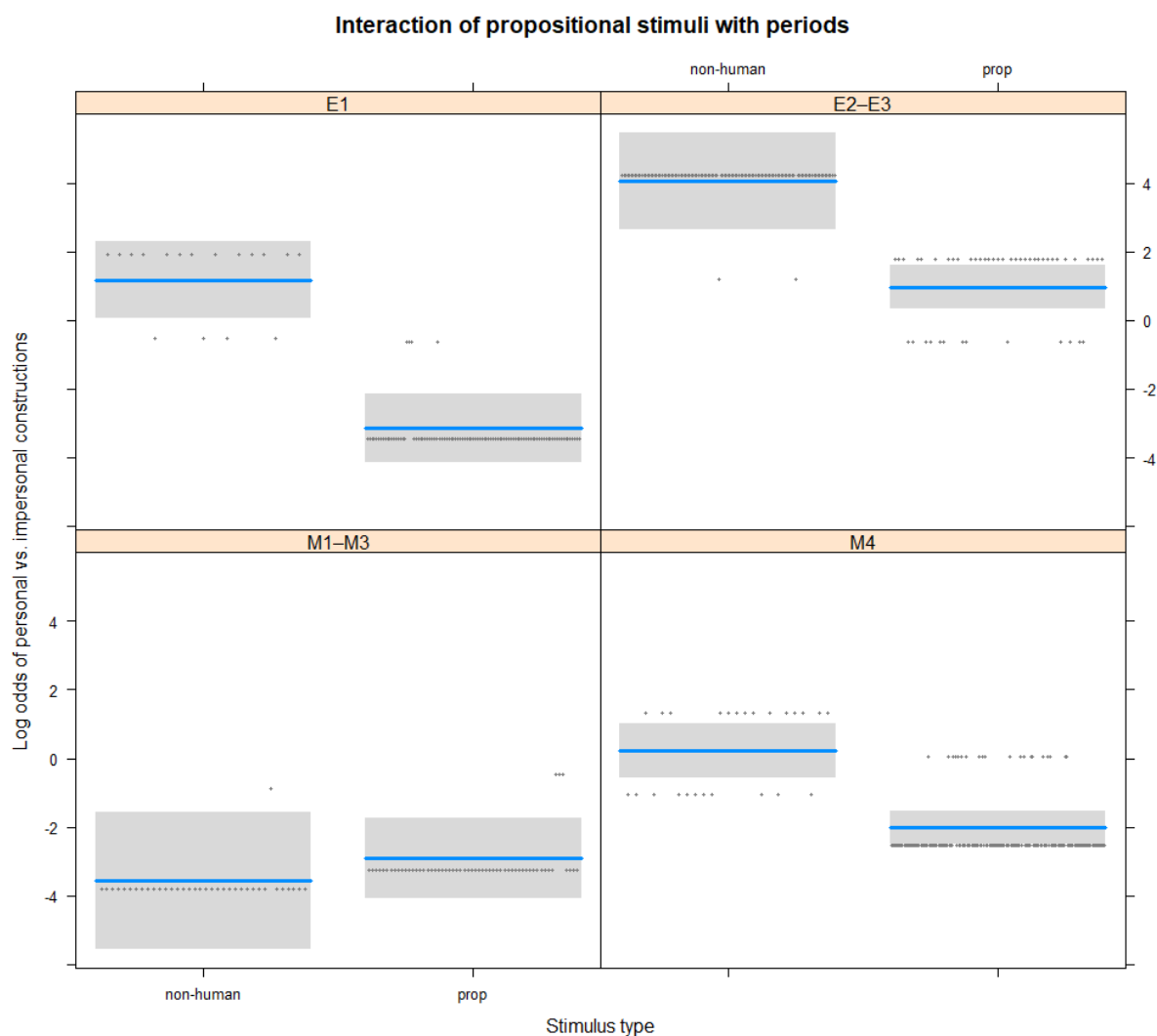


Figure 13 Interaction of propositional vs. non-human stimuli of *liken* with periods.

The situation in the first period M1–M3, which is represented in the bottom-left part of the plot, is similar to the situation discussed for human stimuli in E2–E3. While the associations of propositional and non-propositional stimuli in this period are not significantly different, the observed increase in the log odds of personal constructions with propositional stimuli by a coefficient of 0.6466 ($p = 0.58$) is the reverse of all other periods, in which non-propositional stimuli, which are restricted to non-human stimuli in this version of the model, consistently increase the log odds of personal constructions. The lack of significance in the first period M1–M3 is understandable, since personal constructions only begin to emerge in the data for *liken* during this period, and the contrast between non-propositional stimuli (2.8% personal) and propositional stimuli (5.2% personal) rests on a very low number of observations of personal constructions (1 and 3 instances respectively). The other periods exhibit a robust effect, however. In all of them, the log odds of personal constructions are significantly lower for propositional stimuli than for non-propositional stimuli. The top-right part of the plot with the final period E2–E3 also illustrates that the overall effect is sustained longer (log odds -3.0904,

$p < 0.001$) than the equivalent effect observed in that period for human vs. non-human stimuli, which contrasted less strongly and not significantly. This is specifically due to a continued attestation of impersonal constructions with propositional stimuli in E2 (13 instances or 31.7% of all instances of propositional stimuli in that period), which exceeds that of the other types of stimuli, but subsides completely in the last period E3 (0 instances). Finally, the top-left of the plot indicates that the effect is particularly strong in E1 (log odds -4.3250, $p < 0.001$) compared to both the preceding and the subsequent period, although it differs significantly only from the former, in which the negative log odds are raised by a log odds ratio of 2.0756 ($p < 0.05$), while the latter difference falls short of significance (log odds ratio 1.2346, $p = 0.26$). The coefficients and significance values were obtained from the same model underlying the plot with E1 as the reference period.

The apparent reversal of the effects observed for human vs. non-human stimuli in the final period E2–E3 and for propositional vs. non-propositional stimuli in the initial period M1–M3 clearly contrasts with the overall effects observed for these variables across all periods, but, at the same time, the shift in the odds of the two construction types from exclusively impersonal to virtually exclusively personal constructions implies little potential for significant differences in the marginal periods. The small and insignificant effects observed in these periods are thus in keeping with the nature of the diachronic change that the regression model represents, although a greater amount of data might serve to substantiate the effects. It is, perhaps, notable that 3 of the 4 earliest instances of personal constructions in the data involve propositional stimuli, although these are, of course, also the most frequently attested type overall. It is also noteworthy that none of these instances contains an expletive. The particularly strong association of propositional stimuli with impersonal constructions in E1 and the continued association in E2 can be related to the sustained use of impersonal constructions that involve with propositional stimuli in conjunction with an expletive, while constructions excluding an expletive are exclusively personal from E1 onwards. Since the apparent reversal of effects in the marginal periods can thus be accounted for by the convergence of the construction types, and since the variability in the strength of the association of propositional stimuli appears to be attributable to the subset of observations that include an expletive, the implementation of the interactions seems less desirable than a focus on the overall effect of the variables. The poorer representation of human stimuli would require a restriction of the interaction of this variable to a binary period division with lowered precision and reduced diachronic control, although an equivalent interaction of propositional vs. non-propositional stimuli with a ternary period division would be possible at the same level of precision. No interactions were therefore

included in the regression model. The preceding discussion has shown, nevertheless, that the distinction of observations with and without an expletive is reflected in the changeable but, at the same time, sustained effect of the association of propositional stimuli with impersonal constructions to a degree that merits its exploration by a dedicated variable. The inclusion of such a variable, which is also suggested by the different distributions of construction types in the respective subsets discussed in section 4.1.1.2.1 above, will serve to control the effect of propositional stimuli for the presence of an expletive.

4.1.3.1.4. Random Effects

As mentioned at the beginning of this section, the observations of *likēn* violate one of the assumptions of logistic regression modelling, since they are not strictly independent of each other but derived from a limited number of corpus files, which represent (samples of) literary texts in the case of the *PPCME2* and collections of letters in the case of the *PCEEC*. The distribution of observations across these divisions of text was described in some detail in section 4.1.1. Although their dispersion is substantial enough, for the most part, to dispel any real concerns about their interdependence, certain texts exhibit a particularly high incidence of *likēn* in standardised and absolute terms, which makes them disproportionately influential in determining the overall database. This becomes even more obvious when looking at the level of individual authors in the correspondence corpus. While the corpus files of the *PPCME2* are probably the best possible representative of the authorial “subjects” from which the samples of *likēn* are drawn, the collections of letters in the *PCEEC* are generally much larger and often contain letters from different authors written across larger timespans. Allowing for this principal difference in the concept of text/author, there are 187 different subjects that contribute at least one observation to the data of *likēn* in the final dataset ($n = 621$). The ratio of total observations and individual subjects suggests that the interdependence of these observations is somewhat balanced by their dispersion across a relatively large number of subjects. The median number of observations per subject is 2. There are several subjects, however, with an incidence of *likēn* that is at least ten times higher, including the sample of *The Book of Margery Kempe* in the *PPCME2* and the letters of William Paget, Dorothy Osborne/Temple and Thomas More. The last of these authors exhibits a particularly frequent use of *likēn*, contributing 60 instances or 9.66% of all observations in the final dataset, but also the letters by Dorothy Osborne/Temple contain a comparatively large number of observations of the verb (35 instances or 5.64%), which thus play an important role in shaping the database.

Since more than 15% of all observations of *likēn* in the final dataset are derived from only two individuals, it seems clear that the personal preferences of these individuals may potentially lead to associations of the predictor variables that are not representative of the total population of language users. There are several obstacles to addressing this issue by the implementation of random effects in the regression model, however. First, the clustering effect is much less pronounced for most of the other individuals, which, in many cases, contribute only 1 or 2 observations each to the dataset. Estimating random intercepts and random slopes for these individuals can only be done with great uncertainty. Second, the observations drawn from the individuals do not necessarily represent all of the different test conditions equally well, since they are not obtained from a designed experiment but rely entirely on the chance of transmission. Third, the individuals are distributed diachronically, which means that their individual coefficients are partly dependent on the general change in the odds of the two construction types, a fact which would either need to be controlled for by the period information or could be used as a controlling variable in its own right. In light of these reservations, the interdependence of observations attributable to their provenance from identical subjects, which constitutes a general problem in most corpus-linguistic studies and affects the present data only in a restricted subset of particularly highly represented authors, cannot be addressed adequately by the implementation of random effects. The profiles of the most influential authors will nevertheless be part of the discussion for which the regression model provides the basis.

4.1.3.2. Interpretation

The following section describes the variables that were included in different versions of the regression model and provides an interpretation of the estimated coefficients. A basic version of the regression model was fitted with construction type as the response variable and three variables as predictors, viz. a diachronic control variable dividing the data into an early (M1–M3), a middle (M4–E1) and a late period (E2–E3), and two conceptual dummy variables representing different semantic types of the stimulus argument. Specifically, these represent human vs. non-human and semantically propositional vs. non-propositional stimuli. The reference level represents the intersection of non-human and non-propositional stimuli, which comprise both (non-human) concrete and (non-propositional) abstract participants. The dimension of animacy was chosen over abstractness for inclusion in the basic version of the model, since it constitutes a conceptually different case compared to concrete stimuli, whose inclusion as a semantic type besides abstract and propositional stimuli results in a variable that

can be conceptualised entirely in terms of degrees of abstractness. After a general evaluation of this model in terms of significance and goodness of fit as well as an interpretation of the estimated coefficients, two substitutions of the predictor variables and the inclusion of one additional control variable will be considered in order to obtain a more comprehensive view of the data. Specifically, these modifications involve the substitution of animacy with concreteness and of semantically propositional stimuli with syntactically propositional stimuli, while the additional variable involves the presence or absence of an expletive.

4.1.3.2.1. Basic Model

The table of coefficients of the logistic regression model in its basic version is reproduced in table 10 below. It includes the system of asterisks commonly used to indicate the level of significance of a given coefficient (***) = very highly significant ($p < 0.001$), ** = highly significant ($p < 0.01$), * = significant ($p < 0.05$), . = marginally significant ($p < 0.1$):

	estimate	SE	p-value	significance
(intercept)	-2.6590	0.6808	< 0.001	***
human	1.2548	0.6300	0.0464	*
proposition	-2.7282	0.3240	< 0.001	***
period_middle	3.1929	0.5954	< 0.001	***
period_late	6.3554	0.6486	< 0.001	***

An indicator of the overall significance of the model and several goodness-of-fit measures can be obtained from the `lm()` function included in the *rms* package in *R*. The first of these is the test statistic of the likelihood-ratio test, which provides a measure of the difference of the fitted model compared to an empty model without predictors and helps to establish whether the former significantly reduces the residual deviance in the data. For the present model, this is the case (LR chi-square 484.56, d.f. 4, $p < 0.0001$). The fact that the model is significantly better than an empty model is, of course, not very surprising, since the diachronic control variable alone accounts for a large share of the observable variation, being designed to represent the change in the odds of the two construction types on the basis of their known distribution. The question is therefore whether the fitted model significantly reduces the deviance compared to a

model that already contains the diachronic control variable. Since the latter model is nested within the former, this can be done in the form of an analysis of deviance using the `anova()` function in *R*. The result of this analysis is that both human and propositional stimuli account for a significant amount of variation in the response variable. In combination, they reduce the residual deviance by 118.39 (2 d.f., $p < 0.001$) to 353.39 on 616 degrees of freedom, although the distinction of the former makes a much more limited contribution to a model that already contains the latter distinction (4.4549, 1 d.f., $p = 0.0348$). The model's concordance index ($C = 0.931$) suggests a very high discriminatory power of the model, but, again, this needs to be compared to the model that already includes the periods as factor. The comparison shows that a high level of discrimination by that model ($C = 0.868$) is improved even further by the fitted model. The model was validated by random resampling with 200 repetitions using the `validate()` function of the *rms* package. This indicated relatively low optimism scores for both the intercept (0.0079) and slopes (0.0189), which suggests that the model does not overfit the data.

Regarding the individual coefficients of the model, the estimated odds of personal and impersonal constructions in the contexts specified by the predictor variables can be verbalised as follows. For the reference period M1–M3, constructions of *likēn* involving non-human non-propositional stimuli are estimated to be impersonal with a percent probability of about 93.5%, or, in other words, the chances of impersonal constructions in this constellation are about 14.28 times higher than those of personal constructions. This interpretation of the intercept is congruent with the fact that the designated reference period involves the phase in which personal constructions of *likēn* are only beginning to emerge in the data. The interesting question is to what extent the different types of stimuli influence the changing odds of the two construction types. According to the model, human stimuli increase the odds of personal constructions by a coefficient of 1.2548, while propositional stimuli decrease the odds of personal constructions by a coefficient of -2.7282 when the control variable is kept constant. Taking the middle period M4–E1 as a reference point and continuing to convert log odds into percent probabilities, this is equivalent to saying that the probability of personal constructions, which is already estimated to exceed that of impersonal constructions in the reference group of non-human non-propositional stimuli ($P = 63.0\%$), is raised to 85.7% with human stimuli and lowered to 10.0% with propositional stimuli in this period. For the late period E2–E3, the respective estimates are 97.6%, 99.3% and 72.5%, indicating that the shift in the odds of the construction types of *likēn* towards personal constructions is expected to be virtually completed by this period, except for instances involving a propositional stimulus argument.

4.1.3.2.2. Modifications

The coefficients in the above regression model can serve as a reference point for several modifications that provide additional information about the associations of the two construction types of *likēn* with the different types of stimuli. First of all, concrete stimuli can be tested as one of the conceptual variables instead of human stimuli. Technically, this implies shifting the subgroup of non-human concrete stimuli from the reference group to the compared group, since human stimuli constitute a special case of concreteness. The estimated log odds ratio of 1.0996 is significant ($p = 0.0228$) but somewhat more moderate than that of human stimuli (1.2548), indicating that non-human concrete stimuli increase the odds of personal constructions less strongly than human stimuli. The fact that the two groups in combination are more strongly associated with personal constructions than the reference group of abstract stimuli shows, nevertheless, that a conceptualisation of the semantic cline in terms of concreteness is a valid supplement to the conceptualisation in terms of animacy. Second, the variable representing propositional stimuli in terms of semantics can be substituted with a variable representing propositional stimuli in terms of their syntactic category. While the actual difference between the two is slight, the former group does include a number of additional cases in which semantically propositional arguments are realised by anaphoric pronouns or PPs. The coefficient of the more narrowly defined syntactic variable is estimated at -2.8396 ($p < 0.001$), which is somewhat more pronounced than that of the semantic variable (-2.7282). In addition, the residual deviance is reduced from 353.39 to 343.18 with 616 degrees of freedom. This indicates that syntactically propositional stimuli have a stronger association with impersonal constructions than propositions according to the more widely applicable semantic definition, and that the inclusion of syntactically non-propositional cases in the reference level is able to account for a greater amount of variation in the response variable.

The third modification involves the addition of a variable that accounts for the presence or absence of an expletive. Two versions of this modification are conceivable, viz. one that represents the distinction across all observations by introducing the appropriate variable to the entire model and one that restricts the distinction to propositional stimuli by introducing the distinction to the appropriate variable. Since only 2 instances of the expletive occur with non-propositional stimuli, the two versions result in very similar coefficients. However, it should be noted that the distinction of human vs. non-human stimuli, despite a stable coefficient, falls just below the significance level (log odds ratio 1.2598, $p = 0.0563$) in the former version, while its effect size and p-value increase slightly in the latter version (log odds ratio 1.3736, $p = 0.03435$). The coefficient of propositional stimuli, on the other hand, is reduced substantially when

controlled for the presence of an expletive. Within the former version of the expanded model, the estimated log odds ratio is lowered to -1.0910 ($p < 0.01$) compared to -2.7282 in the basic model, while the newly introduced distinction of observations with and without an expletive constitutes the strongest influence on the odds of the construction types with an estimated log odds ratio of -4.7010, which is equivalent to a probability of impersonal constructions of about 99.1%, everything else being equal. Propositional stimuli thus continue to exhibit a strong association with impersonal constructions, even if the presence of an expletive is controlled for, but the more pronounced effect estimated by the basic model is clearly attributable to instances that contain an expletive. The inclusion of this distinction also reduces the residual deviance of the model significantly (108.27, 1 d.f., $p < 0.0001$) and raises the discriminatory power ($C = 0.973$), which constitutes a significant improvement. The table of coefficients of the revised regression model is presented in table 11:

	estimate	SE	p-value	significance
(intercept)	-2.8331	0.5652	< 0.001	***
human	1.2598	0.6600	0.05631	.
proposition	-1.0910	0.3708	0.00326	**
expletive	-4.7010	0.6778	< 0.001	***
period_middle	3.4027	0.5792	< 0.001	***
period_late	7.1251	0.7577	< 0.001	***

The optimism scores obtained for this model by executing 200 repetitions of random resampling with the `validate()` function suggest that overfitting of the data does not present a serious concern (intercept = -0.0007, slopes = 0.0360). Despite the marginal significance of the distinction of human and non-human stimuli, the variable was retained together with the other variables in the majority of instances of stepwise backward elimination performed by the same function (135 of 200 times). The revised model is thus superior to the basic model in many respects, and it also highlights the impact of the theoretical decision to subsume observations with oblique experiencers and observations conforming to particular types of word order with an expletive under a single construct of impersonal constructions. The decision can be motivated by the oblique case marking that is discernible with experiencers in both contexts whenever it is overtly marked, but the question remains whether the strong association of

impersonal constructions and the subtype with an expletive should be regarded as part of the main effect established for the different types of stimuli or as a separate effect which the main effect needs to be controlled for. The discussion of this issue and of the results of the regression model in general will be the focus of the next section.

4.1.3.3. Discussion

The diachronic distribution of the different construction types of *liken*, which was presented in section 4.1.2.1 and adapted as control variable in a logistic regression model in section 4.1.3.1.1, displays a distinct shift from impersonal to personal constructions. No instances of personal constructions are attested within the present dataset for the earliest period M1 (1150–1250), and only sporadic instances are attested between M2–M3 (1250–1420). The middle period M4–E1 (1420–1570), by contrast, displays a high degree of variation between personal and impersonal constructions. It thus carries the most weight in assessing the influence of different semantic types of stimuli in the diachronic change. Near-completion of the change is attained during the penultimate period E2 (1570–1640), during which impersonal constructions decrease significantly in proportion with all types of stimuli except propositions, and they become generally obsolete in the final period E3 (1640–1710). The development of *liken* can thus be characterised as that of an impersonal type-ii verb into a personal type-iii verb, even though the large majority of stimulus arguments, propositional or non-propositional, are not overtly marked by nominative case in the early periods M1–M3.

Regarding the influence of semantic factors on this development, the regression model indicates significant associations with the construction type for most of the semantic distinctions that were made. Assuming that such associations in a succession of discrete synchronic periods can be interpreted as promoting or demoting factors in a scenario of diachronic change, the semantic distinctions are thus clearly relevant to the research question initially posed. The most pronounced association that emerges from the data is that of propositional stimuli and impersonal constructions, although the strength of this effect only applies when the presence of an expletive is not controlled for. If observations with and without an expletive are regarded as a single semantic group, the presence of propositional stimuli constitutes a significant retarding factor in the shift from impersonal to personal constructions. Human and concrete stimuli, on the other hand, exhibit significant or near-significant associations with personal constructions when compared to non-human and abstract stimuli respectively. In terms of their impact on the diachronic change under discussion, these types of

stimuli can thus be considered contexts that advance the spread of personal constructions with nominative experiencer arguments. While not all groups within the more fine-grained distinction of human, non-human concrete, and abstract stimuli can be demonstrated to differ significantly in their association with the construction type, the semantic cline that emerges from the estimations of their respective effect sizes is conceptually plausible and diachronically stable, despite the observation of a relative increase in the association of concrete stimuli with personal constructions, which can be related to their more frequent attestation in the later periods compared to human stimuli.

It is obvious that the associations estimated by the regression model are in stark contrast to the hypothesis developed and formulated at the outset of this study. According to the tested hypothesis, the use of the innovative construction involving nominative experiencer arguments is expected to occur earlier or more frequently in contexts in which the experiencer argument can be interpreted as agentive in comparison with the stimulus argument. Since experiencer arguments are relatively constant in their semantic properties, and since individual inspections of the surrounding context could lead to subjective and potentially circular conclusions, the establishment of the conditions for this interpretation rests entirely on the indirect evidence provided by the semantic properties of the stimulus argument. According to the operationalisation of agency discussed in section 2.2.2.3, animate and concrete stimuli were expected to favour an agentive interpretation of the stimulus argument and therefore to disfavour an agentive interpretation of the experiencer argument, while non-human and abstract stimuli were expected to disfavour an agentive interpretation of the stimulus argument and therefore to favour an agentive interpretation of the experiencer argument. The higher potential for an agentive interpretation of the experiencer argument in the context of non-human and abstract stimuli should result in the advanced spread of nominative case marking, but the opposite is the case according to the data. While the earliest attestations of personal constructions are, in fact, either abstract or propositional and thus seem to conform to the hypothesis, the significance of this fact is severely qualified by the insufficient amount of data, the requirements for which are particularly high during the early and late periods of limited contrast. With regard to the overall effect of the different types of stimuli estimated across all periods, the opposite tendency is clearly prevalent.

Apart from the apparent reversal of the observed effect compared to that predicted by the hypothesis, there are a number of additional points that require discussion. First of all, the difference in the strength of the associations of semantically and syntactically propositional stimuli is noticeable. While syntactically propositional stimuli perform somewhat better in

terms of the amount of variation they are able to explain, there are still good reasons for considering the slightly wider definition of semantically propositional stimuli a more suitable predictor. As mentioned before, the categorial realisation of propositional stimuli by clausal constituents can be regarded as a correlate of the semantic properties of their referents. Also, the majority of lexical and syntactic categories distinguished for the purposes of annotation did not exhibit any straightforward associations with the construction types. The fact that the only category for which a significant association could be ascertained is largely coextensive with a semantic category is thus probably not coincidental, even in spite of the fact that the conflation of finite and non-finite subordinate clauses into a single category of propositions was based on their semantic equivalence. Alternatively, the distinct association of syntactic propositions in comparison to other the categories could be attributed to their phonological weight, which is typically greater than that of the other types of constituents. It should be remembered, however, that the set of observations included in this category contains a substantial number of inferred stimuli without phonological expression. Unless the concept is defined in a way to include a measure of the structural complexity of unexpressed constituents, which might play a role in terms of processing difficulty, weight does not appear to be demonstrable as a crucial factor underlying the association of propositional arguments. It should also be noted that personal pronouns and non-personal pronouns, which would be expected to behave similarly based on the criterion of phonological weight, exhibit different associations with the construction type, aligning them with the semantic properties of their referents. The question remains why observations including semantically propositional arguments should display a weaker association with impersonal constructions than the slightly more restricted subset of observations realised by clausal constituents if semantic complexity of the referent is, indeed, the decisive criterion.

The answer to this question involves the subset of observations that contain an expletive constituent, which is typically realised by pronominal (*h*)*it* in the present data. As the revised logistic regression model has shown, the association of observations including propositional stimuli, to which the presence of an expletive applies foremost, is greatly strengthened by the presence of an expletive. The fact that propositions are always realised by clausal constituents in this context, either expressed or inferred, but do not include semantically propositional arguments realised by other categories can account for the observable difference between the two groups. The association of observations with an expletive and impersonal constructions is so strong that it necessitates the question of whether their inclusion in a model of variable choice between two construction types is, in fact, justified or whether *it*-constructions need to be

considered a distinct construction type. At the same time, the constituent order in *it*-constructions was used as one of the criteria for the disambiguation of the construction types discussed in section 3.4.1.2.1. If the presence of an expletive presents a definitional criterion of impersonal constructions, its inclusion as a predictor in terms of a main effect would be circular, and the same would be true of the distinction of propositions involving an expletive as a separate morphosyntactic predictor.

There are at least two points, however, that argue in favour of the inclusion of observations with an expletive within the general concept of a constructional choice between personal and impersonal constructions. First of all, a small number of personal observations do exist which involve an expletive or, in alternative terms, a cataphoric pronoun in object position that precedes a clausal constituent within the same clause and is thus superficially equivalent to an expletive. This means that the construction type is strictly speaking independent of the presence of an expletive, despite the very strong association of the latter with impersonal constructions. The 4 observations in the data for *liken* that contain such a constituent in object position are probably too numerous to be simply discounted as errors, and the minimally restricted theoretical framework underlying the present account does not contain any principled reason to exclude the extraposition of object constituents as a grammatically viable option. Second, the nominative case marking of the experiencer argument in these instances participates in a contrast that extends equally to observations with and without an expletive. Since the instances of observations with an expletive that conform to constituent-order types in which the experiencer argument occupies a prototypical object position uniformly exhibit oblique case marking of the experiencer argument whenever this is overtly manifested, their inclusion under a single concept of impersonal realisation seems justified according to this criterion. The importance of expletives for the disambiguation of the construction types, which is necessarily tied to the constituent order that their presence entails, is thus limited to the subset of instances for which covert oblique case marking needs to be assumed. Given the increasingly limited functionality of morphological case marking with regard to nouns and the second-person singular and plural pronoun *you*, the evolvement of the construction of impersonal verbs from a morphological choice with relative syntactic freedom into a morphosyntactic choice with restricted syntactic freedom is congruent with the general development observed for English grammar during the period under investigation. For the language user, the employment of extraposed propositional arguments in combination with an expletive thus opens up the possibility of maintaining the interpretation of experiencer arguments as an affected entity in contradistinction to the more agentive interpretation which, by virtue of the relative type and

token frequency of verbs that include an agent in their argument structure, constitutes the prototypical semantic role mapped onto subjects in canonical active clauses. The retention of this option results in a considerable delay of the eventual demise of impersonal constructions with *liken*.

The integration of instances with an expletive in the concept of a binary contrast between personal and impersonal constructions thus appears justifiable, but a noticeable concentration of these constructions and related instances involving propositional stimuli without expletive is nevertheless observable, both in terms of their diachronic distribution and in terms of their communicative function. The distribution of the different types of stimulus arguments of *liken*, which was already presented for the restricted subset in section 4.1.2.2, exhibits a marked concentration of observations with propositional stimuli in the middle period (M4–E1) and, to a somewhat lesser degree, in the preceding period E3. While the other types of stimuli are generally much less frequent in comparison to propositions, human stimuli also exhibit a particularly skewed diachronic distribution when considered in isolation. Within the full dataset of *liken* that includes non-finite instances and passive clauses, ambiguous constructions and cases of coordinated human and non-human stimuli (n = 831), 67 or about 79.8% of all 84 instances of human stimuli are attested in the EModE periods E1–E3 and 57 instances or about 67.9% in the two final periods E2–E3 alone, while the ME periods M1–M4 only account for 17 or about 20.2% of all instances. This distribution is, of course, partly related to the fact that the EModE periods account for more than half (53.6%) of the total corpus data, but the emerging trend of a diachronic increase in the occurrence of human stimuli with *liken* is also observable when the frequencies are considered in relation to the number of corpus words in each period. Table 12 below presents the absolute and standardised frequencies of human stimulus arguments of *liken* per 100,000 words:

	M1	M2	M3	M4	E1	E2	E3
absolute	7	0	2	8	10	27	30
standardised	2.71	0	0.47	1.05	3.23	2.96	5.40

While it is difficult to judge the significance of the absence of human stimuli in M2 given the small sample size in that period (93,999 words), the low frequencies in M3 and M4 are based on larger samples and contrast substantially with the later periods E1–E3. With 2.96

occurrences per 100,000 words compared to 3.23 occurrences per 100,000 words, the standardised frequency of *liken* is, in fact, somewhat lower in E2 than it is in the preceding period E1, but this does not distract from the general trend. There is, of course, one major exception to this trend, and that is the frequency of human stimuli in the earliest period M1, which is comparable to that of the later periods E1 and E2.

Looking more closely at the instances from M1, it emerges that at least some of them are either directly based on Latin exemplars or written in translational contexts, including two translations of *Mat. 3:17* (CMLAMB1,141.280, CMVICES1,119.1478) and one instance in the context of translations from the *Book of Job* (CMTRINIT,167.2268). While these instances may highlight the general problem of a potential transfer of Latinate features to the indigenous ecclesiastic literature as well as that of the inclusion of multiple citations of the same Biblical reference within the same corpus, it is not immediately apparent how this would disqualify them as genuine instances of *liken* with human stimuli, since the authors arguably had a lexical choice when using the verb in translation, even if this choice was limited by an existing tradition. The impression that the relevant instances reflect the vernacular option of using human stimulus arguments with *liken* during the early ME period is supported by the fact that some of the other instances present additional indications of a causative interpretation of the verb, which is in agreement with the hypothesis about human stimuli formulated at the outset of this study. In these instances, *liken* occurs with volitional *willen* ‘to want’ (CMTRINIT,29.386) and as part of an adverbial clause of purpose *for to liken* ‘in order to please’ (CMTRINIT,215.3002), both of which cases presuppose a causative interpretation of the verb. This interpretation does not appear to carry over to the later occurrences of human stimuli, however, since these consistently exhibit the receptive interpretation of the verb, which is most obvious in the many passive constructions encountered in E2, and since they generally do not result in the hypothesised influence on the construction type.

The diachronic differences in the representation of different types of stimuli also need to be viewed in relation to the different types of genre represented by the two corpora. Abstract stimuli, for example, which are generally distributed more evenly than the other types, constitute the most frequently attested type of stimulus argument in the early periods M1 and M2, but this observation is clearly related to the fact that abstract moral features constitute a prominent subject matter of the early English literature represented by the *PPCME2*. The changes in the frequency of human and propositional stimuli, on the other hand, largely take place within a single corpus, viz. the *PCEEC*. As such, they are about as representative of changes in the general language of the time as the sampled correspondence possibly can be.

The typical subject matter of this correspondence may, of course, change over time and thus potentially contribute to the greater or lesser prominence of certain semantic types of stimulus arguments, but it is equally possible to interpret the distribution of the different semantic types of stimulus arguments in terms of a change in lexical properties. Since human stimulus arguments continue to occur during the middle period, such a scenario would presume a bifurcation of *likem* into two separate lexical entries, one containing the impersonal construction with propositional arguments and an optional expletive in a specialised communicative function, and the other containing personal constructions including human and other types of stimuli in a more general function. The former entry of *likem* eventually becomes obsolete and is superseded by the latter, which accounts for the absolute and relative increase in human and concrete stimuli during the late period. The described clustering of the different types of semantic stimuli with the constructions types and with the different periods is visualised in figure 14 in the form of a factor map, which was generated with the help of the `MCA()` function in the *FactoMineR* package in *R*:

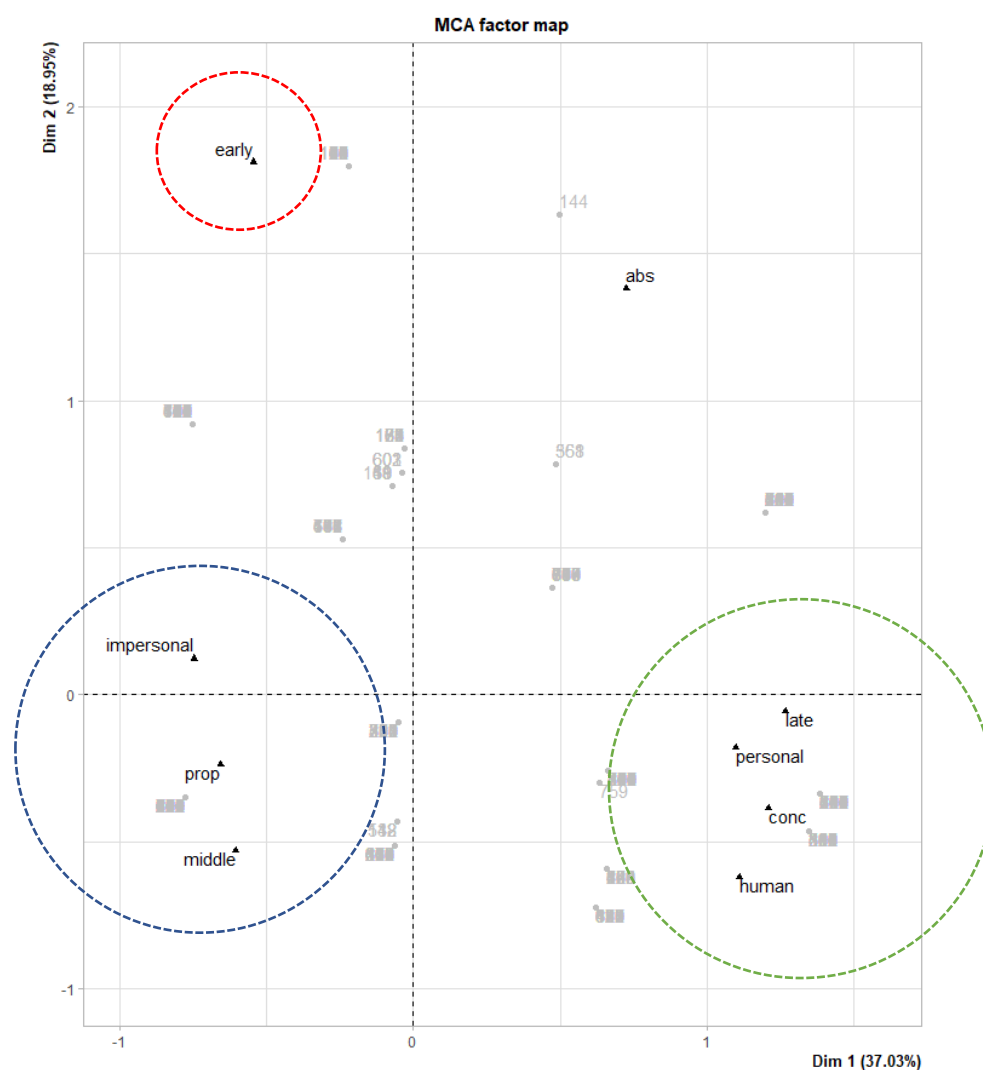


Figure 14 Factor map of chronological, constructional and semantic features of *likem*.

The two dimensions by which the multiple correspondences between the constructional and semantic feature of *liken* are represented in the above plot conveniently illustrate the proposed bifurcation of *liken* into two separate lexical entries. The main difference between this representation and the regression model discussed above is that the chronological information does not act as a control variable for the distribution of the semantic types in this version. Instead, the associations of the different types of stimulus arguments are considered within the entire dataset, and the chronological sequence that emerges with regard to these associations extends not only to the construction types but also to the complementation by arguments with different semantic features. Whether the resulting clusters are best viewed in terms of separate lexical entries is a matter of theoretical debate that will not be entered into here, but a representation along such lines would at least provide a way of accounting for the apparent skewing in the chronological distribution of the different types of stimuli, which would exceed the mere recognition of different degrees of data availability reflected in the significance levels of the regression coefficients.

The proposed notion of a coexistence of two separate lexical entries of *liken* during the late ME and the early EModE periods finds additional motivation in the functional differences that seem to be attached to the respective clusters of semantico-grammatical features. While personal constructions of *liken* involving non-propositional stimuli appear to denote the experiencer's affection by a sensation of pleasure in a more or less literal fashion, corresponding to the receptive interpretation of the verbal semantics discussed above, impersonal constructions of *liken* involving propositional stimuli and the partly optional presence of an expletive generally convey this literal sense to a much lesser degree, which is tantamount a virtual desemanticisation of the verb in the most extreme cases. The desemanticisation of *liken* is most tangible in those instances that were identified as formulaic expressions on account of their recurrent schematic features, which included syntactic and lexical patterns as well as the frequent realisation of the experiencer argument by abstract terms of address. The formulaic nature of these patterns results in an almost inflationary increase in frequency with certain authors, most prominently evidenced by the correspondence of Sir Thomas More. In these contexts, the verb carries notably little of its core meaning as a verb of emotion. Instead, the conventionalised use of *liken* with abstract terms of address in reference to the experiencer and the characterisation of this argument as an oblique constituent by means of morphosyntactically indexed impersonal constructions combine into a circumlocutory way of speech that is highly appropriate for contexts of politeness.

The use of impersonal constructions of *liken* as a feature of polite speech is also indicated by the concomitant use of the hortative subjunctive and desiderative *mouen* ‘may’, which were characteristic of the two formulae discussed above in section 3.4.1.2.2. The first of these primarily serves as a kind of discourse organisation, introducing a new piece of information by means of requesting the addressee’s permission to do so, which constitutes an obvious rhetorical device in the context of written correspondence. The reduced semantic weight of *liken* in these instances is apparent from the fact that the imparted information does not always seem particularly suited to eliciting pleasure on the part of the addressee, unless the act of notification itself is considered pleasurable (cf. e.g. *like it your maistirship to vndirstond bat wyntir and colde weders draweth negh* (PASTON,II,390.450.11597)). In other instances, the request extends beyond the described call for attention and specifically states an action desired of the addressee, whose identity as the performing subject of this action is implicit in the infinitival complement (cf. e.g. *[y]f it might like your Grace when you se tyme conveyent to vse some fauorable wordes to my said Lord Thresorer [...]* (PARKHUR,180.041.707)). The fact that propositional arguments in the data take the form of infinitival clauses more frequently than that of finite subordinate clauses fits well with the observation that *liken*, in these instances, is used with regard to a hypothetical act, whose performance is desired by the speaker, but which cannot, at the time of utterance, be an actual cause of pleasure for the experiencer. The true semantic weight thus lies with the clausal complement and not with the emotion expressed by *liken*, whose function comes close to that of a modal auxiliary verb, since it essentially modifies the initial predicate in terms of modality. Finite subordinate clause can, of course, be used to a similar effect, cf. e.g. *[a]nd yef hit leke you that he shal have hit at this price [...]* (STONOR,II,121.100.1786). In such cases, the hypothetical nature of the statement is frequently indicated by its embedding in a conditional matrix clause.

The alternative use of *liken* is epitomised in the letters by Dorothy Osborne/Temple. Occurrences of the verb are of an equally notable frequency here as the formulaic occurrences in the letters by Thomas More, but they contrast in being of a much more varied kind and typically with reference to more concrete stimuli, including objects and humans, or to more circumscribed abstract notions (cf. e.g. *[y]ou will not bee Jealous though I say I like him [viz. a certain gentleman] very much*, (OSBORNE,177.075.4155) and *but of all the People that ever I saw I doe not like his Carriage towards her* (OSBORNE,175.074.4102)). They are also noteworthy in much more frequently referring to the author’s own feelings or stance of approval. In the final subset of unambiguous finite active clauses, first-person experiencer arguments are more than twice as frequent than second-person experiencers with this author

(22 compared to 9 instances). Incidentally, this is reminiscent of the more “personal” style of writing often ascribed to female writers in sociolinguistic studies. Written in the early- to mid-1650s (E3), the letters by Dorothy Osborne/Temple obviously belong to a later period than the letters by Thomas More, instances from which range between 1519–1534 (E1). As such, their invariably personal construction is less revealing and already characteristic of the type of use of *liken* that continues into Present-Day English. The earlier coexistence of this literal use of *liken* with the much more frequent polite use during the middle periods is indicated by the relatively strong association of human and concrete stimuli with personal constructions, and it is demonstrable, despite a relative scarcity of examples, by instances such as the following, written in 1481: *and I toulde hym of the good whyll that the Whegystons and Dawltons hows [i.e. house] to yow , and how I lykyd the zenge [i.e. young] gentyllwhoman ,* (CELY,107.083.1851)). This instance from M4 combines all the features observed in the letters by Dorothy Osborne/Temple and equally suggests a more literal interpretation of *liken* as a verb of emotion rather than as a conventionalised index of politeness.

Overall, the dichotomy between instances of *liken* with propositional stimuli and those with non-propositional stimuli, mostly notably those referring to human and concrete referents, is most deeply entrenched in the data. It is observable in the vastly different coefficients of the regression model as well as in the clusters emerging from the multiple correspondence analysis, in which diachronicity is not controlled for. While propositional instances frequently co-occur with an expletive and display an even stronger association with impersonal constructions in this case, the finding that such instances constitute a clearly distinguishable type of *liken* in both formal and functional terms that can be construed as a separate lexical entry is not a circular result of the inclusion of so-called *it*-constructions in the data, since the presence of an expletive is not a necessary condition on the observed association and since the morphosyntactic marking of the experiencer argument as an oblique constituent is subsumable under a single concept of a binary contrast between impersonal and personal constructions. Instead, the functional specification as a primarily modifying element in the complex structure of polite speech, which is epitomised in the frequent formulaic occurrences during the middle periods, appears to arise as a natural implicature of the complex semantic structure of propositional arguments. Their syntactic realisation by non-finite complement clauses, in which the identity of the subject of the expressed action and the experiencer argument of *liken* is presupposed, is well-suited to the task of directing polite requests at an addressee, which, in the conventionalised form observed in the sampled correspondence of these periods, may be limited to the organisation of discourse structure by introducing a new topic, but which also serves the

wider function of expressing a range of different requests. The extent to which this function is exclusive to the semi-private and official correspondence, particularly the supplicant letters addressed to various dignitaries, which constitute a large share of the data for the later periods, is difficult to judge without the comparison of a corpus representing alternative genres. At any rate, the personal use of *liken* with mainly human and other concrete stimuli that becomes dominant towards the end of the period under investigation appears to be distinct from the early instances of human stimuli that were discussed for the earliest period M1, since the implication of an agentive interpretation of the stimulus argument, which could be established for some of the early occurrence at least, appears to be absent from the late examples, where the prominent interpretation of the verb remains receptive even when human stimuli are involved. Given the almost complete diachronic separation of *liken* as a modal modifier in contexts of politeness and literal instances of *liken* as a verb of emotion, the question arises whether other lexical items from the semantic field of liking can be identified as complementary in fulfilling the respective functions. Potential candidates will be discussed in the following section.

4.2. Other Verbs

The selection of verbs for investigation besides ME *liken* was based on two fundamental considerations. On the one hand, the intention was to represent different semantic classes while also facilitating a comparison of semantically related verbs in terms of their correspondence with the research hypothesis. On the other hand, the selection was limited by the employed method of identifying pertinent verbal tokens in the corpora, since the precision of the retrieval of these tokens was not equally potent for all potential candidates but sensitive to the number of spelling variants and existing homophones or homographs of a given lexeme. Thus, while a verb of cognition like ME *thinken* would have been an interesting candidate on account of its representation of a separate semantic class compared to verbs of emotion, the high number of spelling variants recorded for this verb in the *MED* and the potential confusion of *thinken* ‘to think’ < OE *þencan* and *thinken* ‘to seem’ < OE *þyncan* made its inclusion impractical. For a verb like ME *lusten* ‘to desire’, an inclusion of the attested spelling variants with the root vowels *o* and *i* would lead to a partial overlap with participial and past-tense forms of ME *losen* ‘to lose’ as well as a general overlap with the homophone verb ME *listen* ‘to listen’. The manual exclusion of false-positive hits resulting from these overlaps may not be too time-consuming for the more obvious cases, but if the expected size of the resulting dataset was considered to be too small to be informative, this procedure was not undertaken. Another case is the verb of

cognition ME *meten* ‘to dream’, whose forms were extracted from the corpora but not annotated because the majority of a sample of the extracted forms were identifiable as forms of ME *meten* ‘to meet’ or, in some cases, of ME *meten* ‘to measure’, leaving too few forms of the verb of interest for investigation. Similarly, the number of potential forms of ME *dremen* ‘to dream’ was too low to serve as a representative basis so that, ultimately, none of the verbs of cognition initially considered were included in the study.

The verbs whose data were fully processed fall into three distinct semantic groups, viz. near-synonyms of *liken* ‘to please’, including ME *quemmen* ‘to please, be acceptable’ and *plesen* ‘to please, satisfy’, other verbs of emotion, which are represented by *reuen* ‘to regret, feel pity’ and *longen* ‘to desire’, and *wanten* ‘to lack’ as a verb of necessity. Neither *quemmen* nor *plesen* are typical representatives of the process of personalisation of impersonal constructions. The former becomes virtually obsolete during the ME period and the latter adopts the option of an expletive subject while preserving the oblique case of the experiencer argument in most cases. Nevertheless, these verbs are relevant in complementing the information obtained for *liken* and providing additional insights into the division of semantic and communicative functions in this restricted semantic field. The option of comparing verbs which have been discussed by Allen as prototypical examples of distinct grammatical classes based on different identifications of the semantic role of the stimulus argument as either cause or theme motivates the inclusion of *quemmen* in particular. Otherwise, the data for this verb are comparatively sparse and largely restricted to the earliest period of Middle English. The other two verbs of emotion, *reuen* and *longen*, are not represented by large numbers of instances in the two corpora either, and while the search for tokens of *wanten* yields a more substantial number of observations, these frequently do not fall into the subset of finite active clauses with two explicit arguments to which the analysis of *liken* was restricted. This means that, effectively, none of the verbs that were investigated in addition to *liken* qualify for a quantitative analysis in terms of a logistic regression model. Instead, the data and distribution of variable levels will be presented and discussed for these verbs along similar lines as those of *liken*, but the limitation of their quantitative representation in general and of the attestation of different variable levels in particular require all references to the observed frequencies to be understood without the implication statistical significance. Any inferences based on such data will need to be developed from a qualitative discussion of their prominent features, but the additional verbs nevertheless provide several interesting points of comparison with *liken* and thus help to inform the general picture that emerges about the development of oblique-experiencer verbs.

4.2.1. Quemen

The verb ME *quemen* ‘to please’ < OE *cweman* occurs alongside a variant ME *iquemen* ‘id.’, although it is most frequently attested without the verbal prefix *i-* in Middle English. Since this prefix generally becomes obsolete for English verbs, *quemen* may be thought to at least partly continue OE *gecweman* ‘id.’ as well, even though separate forms including the prefix are also attested in Middle English. The verb is discussed by Allen as prime example of a Dative Object verb, whose experiencer argument does not acquire behavioural subject properties and whose second argument is identified as a cause argument rather than a theme argument on account of several features, including its frequent reference to animate participants, its realisation by pronominal constituents and its preposed position relative to the experiencer argument. While the glossing of the verb by PDE *please* motivates its inclusion in the group of near-synonyms of *liken* or, more generally, in the class of verbs of emotion denoting pleasure or desire, it is worth noting that its scope of meaning is not identical to either that of *liken* or *pleasen*, as notions including satisfaction or gratification, being serviceable or compliant, and also being suitable or appropriate are recognised and recorded by the *MED*. In the last sense, *quemen* thus borders on *behove*-verbs like *behoven* and *semen*. Although the verb becomes largely obsolete during the ME period, instances of PDE *queem* ‘to fit closely, to smooth out’ are recorded in the 20th century for Scottish English, albeit in citations that indicate rare or poetic usage (cf. *OED* s.v. *queem*). The verbal argument corresponding to the experiencer in these instances takes the form of a direct object, which means that no personalisation of the verb by extension of nominative case to a formerly oblique constituent has taken place. The main question about *quemen* is thus whether its attestations after the OE period investigated by Allen continue the typical reference of the stimulus argument to animate participants and the realisation of this argument by pronominal constituents.

4.2.1.1. Database

The full dataset of *quemen* contains only 87 observations, all of which were obtained from the *PPCME2*, while an identical query in the *PCEEC* yielded no results. It is worth noting that this count includes 6 forms that contain the prefix *i-* and that such forms are assigned to a separate lemma *iquemen* (MED23304) in the *MED*. The limited attestation of the compound verb compared to the simple verb is the reverse of what Allen found in her corpus of Old English, where *gecweman* was several times more frequent than the simple verb *cweman* (cf. Allen 1995: 144–149 and section 2.2.1.2.1 above). Given the earlier predominance of the compound verb

and the general abandonment of the prefix *i-* in Middle English, it seems plausible to regard ME *quemen* as a successor to both the simple and the compound verb, even if the compound is retained in certain ME manuscripts like the *Lambeth Homilies*, which contains 4 of the 6 prefixed forms. This manuscript may also be considered more archaic, since the homilies it contains are essentially transliterated versions of 11th-century texts. Judging by the few overall attestations of *quemen* and *iquemen* in this collection of homilies, their use is largely equivalent. Since an individual investigation of the few prefixed forms would not be particularly revealing, they were simply retained in a single dataset alongside *quemen*.

The most conspicuous fact about the data for *quemen* and *iquemen* is that 61 or about 70.1% of the total 87 observations derive from a single text, namely the 12th-century homiletic poem *Ormulum*. Since the *PPCME2*'s sample of this work contains only 50,579 words, this indicates an incidence of about 120.6 occurrences per 100,000 words, which is high even when compared against the frequent formulaic instances of *liken* in some of the collections of the *PCEEC*. The two texts with the second- and third-highest number of observations are the *Trinity Homilies* and the aforementioned *Lambeth Homilies* with 9 and 6 observations respectively, which correspond to about 21.5 and 28.9 instances per 100,000 words in each of the samples. These figures not only illustrate the relative dominance of the *Ormulum* on the total data, but they also suggest that the use of *quemen* is rather specialised and largely restricted to homiletic literature, which is, of course, particularly prominent in the early parts of the *PPCME2* in general. Nevertheless, the general absence of *quemen* from the later and slightly more varied sections of the corpus is notable. The single exception to this absence is presented by 3 instances in the *Ayenbite of Inwyte*, a translation of the French original *Somme le Roi* dating to 1340, which falls into the ME period M2. All other observations derive from the earliest period M1. This narrow distribution of *quemen*, both in terms of diachrony and genre, makes it more difficult to judge the representativeness of the extracted observations compared to the general use of this verb at the time of its attestation. Since the citations in the *MED* suggest a wider distribution of the verb across different text types, it needs to be assumed that the extracted dataset only represents part of this use.

Regarding the clause type of the observations in the dataset of *quemen*, the first point to note is that 28 or about 32.2% of all instances take the form of infinitival clauses, which frequently function as complements to adjectives, cf. e.g. *zeorrrnfull Crist to cwemenn* 'eager to please Christ' (CMORM,I,54.518) and *tor to cwemenn* 'difficult to please' (CMORM,I,248.2017), or as adverbial clauses of purpose, cf. e.g. *for to quemen ure drihten* 'in order to please our Lord' (CMLAMBX1,137.1396). This proportion of non-finite clauses,

which are exclusively infinitival, is much higher than that observed for *liken*, where 6.9% of all observations were non-finite, including all types of non-finite clauses. Specifically, the prominent use of infinitival forms of *quemen* as purpose clauses seems to be connected to the homiletic subject matter, which frequently revolves around the discussion of righteous behaviour according to the law of God and includes repeated admonitions of the reader to behave according to God's will. Passive constructions, on the other hand, are not very prominent, with only a single instance identified in the corpus, cf. *þe king wes swiðe icwemet* 'the king was very pleased' (CMKATHE,27.131). The interpretation of the verb in both the non-finite clauses and the passive instances is consistently causative, which means that the PRO subject of the infinitival clause bears the semantic role of the stimulus and the subject of the passive construction bears the role of experiencer. This accords with the early non-finite instances of *liken*, but it contrasts with most later instances of *liken* and all of the relatively late passive instances, whose interpretation is receptive. The retention of the stimulus argument as the conceptual starting point of the predicate *quemen* coincides with the preservation of oblique case marking of the experiencer argument whenever this is overtly marked. Since an experiencer argument is typically present in the infinitival clauses, where only the stimulus takes the form of a null constituent, it is potentially analysable in terms of its case marking, which serves as the main criterion for the analysis of the construction type. The following presentation of the construction of *quemen* will nevertheless be restricted to finite active clauses, following the procedure for *liken*. The description of the semantic properties of the stimulus argument and its realisation by different lexical and syntactic categories, on the other hand, will be based on the full dataset of observations containing two arguments, since it is not primarily intended as a characterisation of the database underlying a regression model but as a general characterisation of the observable lexical properties of *quemen*.

4.2.1.2. Variables

4.2.1.2.1. Construction Type

The construction of the 58 finite active observations of *quemen* that remain after the exclusion of non-finite instances and passive clauses was identified as impersonal in all cases in which an analysis of the construction type was possible according to the criteria discussed in section 3.4.1. Specifically, 47 observations were analysed as impersonal constructions, while 11 observations remained ambiguous, typically because the experiencer argument was realised by a morphologically ambiguous NP and no other criteria for the identification of the construction

type were applicable. Morphological ambiguity of the experiencer argument was a more widespread phenomenon, however, since only 21 instances could be analysed directly based on their case marking. An additional 21 instances were categorised as impersonal constructions based on the case-identity constraint, according to which the experiencer argument and the stimulus argument cannot receive identical case marking so that the covert case marking of the experiencer argument can be established whenever the stimulus argument is overtly marked. Since the stimulus argument of *quemen* was frequently realised by pronominal constituents marked by nominative case, this facilitated the identification of the construction type as impersonal in these cases. 4 additional instances were analysed as impersonal constructions based on agreement of the verbal form with the stimulus argument. Given that almost all observations derive from a single period and predominantly from a single text, it seems less likely that the disambiguation of additional instances or the availability of further observations from the same texts would change the resulting classification of *quemen* as an impersonal verb with consistent oblique case marking of the experiencer.

4.2.1.2.2. Semantic and Syntactic Properties

The semantic attributes of the stimulus argument of *quemen* are consistent with an interpretation of this participant as potentially agentive. After the exclusion of 3 observations in which the stimulus argument was not explicitly specified and also not inferable from the context but of a more general nature, 78 or about 92.9% of the remaining 84 observations involved a human stimulus argument. This is an overwhelming majority, which contrasts distinctly with the much more limited use of human stimuli with *liken*, although a certain number of comparable instances were recorded in the earliest period M1 for this verb as well. The preponderance of human stimuli with *quemen* is even greater if 2 instances that refer to the soul as stimulus argument are categorised as human. Specifically, these are *monnes saule* ‘man’s soul’ (CMANCRIW-1,II.112.1392) and *Ba gode sawless alle* ‘all the good souls’ (CMORM,DED.L199.46), whose inclusion under the concept of human stimuli seems to be an option, since the soul, even though it is described as a distinct entity in the context of the former citation, may be interpreted as a metonymical reference to humans in the spiritual literature. For the sake of consistency with *liken*, both instances were nevertheless counted as non-human, even though they behave similarly to other human stimuli of *quemen* and do not necessarily form a separate category. The 4 remaining non-human instances refer to abstract notions, cf. e.g. *mazzdennhad & widdwesshad* | | & *weddlac* ‘maidenhood, widowhood and wedlock’

(CMORM,I,160.1313), which are more clearly distinct from human stimuli. There are thus essentially two main types of stimuli of *quemen* in terms of animacy and abstractness, viz. human stimuli and, to a much lesser extent, abstract stimuli. Non-human concrete stimuli do not play an important role as arguments of *quemen*, apart from the 2 debatable instances of human souls, and also propositional stimuli are notably absent from the data. This, again, contrasts with the attestation of *liken*, which includes propositional stimuli also in the earliest period M1, even though these are by far not as frequent with *liken* at this point as in the late ME and the early EModE period.

Regarding the categorial realisation of the arguments of *quemen*, another point in Allen's finding is confirmed by the data, which is the frequent realisation of the stimulus argument by pronominal forms. Out of 84 instances with an expressed stimulus, 21 take the form of a personal pronoun and another 27 instances are realised by other pronouns, mostly relative or anaphoric pronouns. This high incidence of pronominal forms in general and of personal pronouns in particular can, of course, be attributed to the fact that the stimulus argument frequently refers to human participants, which have a higher chance of being the topic of discourse and thus of qualifying for pronominal realisation on account of their givenness. A greater tendency towards givenness of the stimulus argument is also suggested by the fact that it is realised by a null constituent in 33 additional instances, either by PRO in infinitival clauses or deleted under coordination, while only 3 remaining instances take the form of a full NP. The experiencer argument, on the other hand, which is specified in all of the 87 observations, is realised by a full NP in 58 instances and by personal pronouns in 23 instances, while null constituents and other pronouns comprise only 3 instances and 1 instance respectively. The greater tendency of the stimulus argument of *quemen* towards pronominal realisation compared to the experiencer argument is the opposite of the situation with *liken*, where the experiencer argument was much more frequently realised by personal pronouns than the stimulus argument, even in the early period M1. This suggests that the stimulus argument is generally the more topical element and the better conceptual starting point for propositions involving *quemen*. At the same time, the semantic range of the experiencer in the attested examples is restricted to such an extent that it was probably anticipated by the reader. Variations of *Godd* 'God', *Crist* 'Christ' and *Drihhtin* 'the Lord' account for 51 of the 58 full NPs, and most of the pronominal instances equally refer to either of these or to *Haliz Gast* 'the Holy Ghost'. In the context of the homiletic literature to which most occurrences of *quemen* in the corpus are restricted, the experiencer is thus similarly given as the stimulus argument, even though this givenness results from the general subject matter, while the givenness of the stimulus argument can vary based

on the topical participant in a given context. Finally, 2 out of 3 instances of experiencer arguments in the *Ayenbite*, which is chronologically later than the other observations, are realised by PPs rather than oblique or morphologically ambiguous constituents, a fact which Trips/Stein (2019: 249) interpret as a transfer of contact-induced copying of syntactic constructions from French via the loan verb *plesen* to the indigenous verb *quemen*.

4.2.1.3. Discussion

The data from the *PPCME2* corroborate Allen's findings about the animacy and the frequent pronominal realisation of the stimulus argument of *quemen* also for the early ME period. The fact that virtually all of the data that were retrieved fall into M1 prevents a more detailed diachronic account, however. Judging by later citations from the *MED*, not only the distribution of *quemen* across different literary genres was wider than the present sample suggests, but also the animacy of the stimulus argument was more variable outside of the homiletic contexts that are most prominent in the early section of the corpus. This makes it difficult to judge whether the observed features of *quemen* should be regarded as lexically specified features or whether the prominent use of animate stimulus arguments is an epiphenomenon of the narrow usage that prevails in the data. An additional point of Allen's findings that is confirmed by the data and which sets *quemen* apart from *liken* is the relative constituent order of the experiencer argument and the stimulus argument. Only a single observation among 50 finite active clauses with two overt constituents exhibits a constituent order in which the experiencer precedes the stimulus argument, and this may be an instance of contrastive word order, cf. *to zome / ha wyle queme [...] To opren / ha wyle harmy*, 'some he wants to please, others he wants to harm' (CMAYENBI,23.342–344). In all other cases, the nominative stimulus argument precedes the oblique experiencer argument, which means that *quemen* essentially exhibits the typical syntactic properties of a transitive verb already during the early ME period.

The traditional notion that the stimulus argument of *quemen* generally represents a more volitional participant than the equivalent argument of *liken* can be related to several additional factors besides its animacy and frequent pronominal realisation. First, the use of *quemen* as the predicator of non-finite purpose clauses suggests that the experiencer's disposition towards the stimulus can be controlled by actions that the stimulus performs. This fact is also apparent from the repeated specification of instrumental adjuncts such as *Purh heore rihhtwisnesse* 'by their righteousness' (CMORM,I,10.206) and *Wipþ spellless & wipþ dedess* 'with words and with deeds' (CMORM,I,278.2273), which explicitly state the means by which the experiencer is

supposed to be pleased. Another indication is the use of *quemen* with the (pre-)modal verbs *willen* ‘to want’ and *mouen* ‘may’, which indicate volition and ability of the stimulus (on the combination of *quemen* with *willen* cf. also Trips/Stein 2019: 256). A specific formulation that occurs several times in slight variations in the *Ormulum* and beyond is *Hu þu mihht Drihhtin cwemenn* ‘how you may please the Lord’ (cf. e.g. CMORM,I,40.420), which, again, illustrates the prominence of a devout and virtuous way of life as a subject matter in the early ME sources, and which highlights the stimulus’ ability to intentionally pursue such a course. The more frequent implication of volition of the stimulus argument of *quemen* compared to *liken* may thus partly result from the literary contexts in which the verb is attested. At the same time, the fact that *quemen* appears to be the default verb in these contexts speaks in favour of its lexical compatibility. Overall, the attributes of *quemen* are consistent with the diachronic hypothesis that stimulus arguments with a greater potential for an agentive interpretation disfavour the extension of nominative case to the experiencer argument, but given the lack of diachronic change in the construction of *quemen*, this is not really demonstrable. The failure of *quemen* to evolve into a personal verb with a nominative experiencer throughout the later periods recorded in the *MED* may, of course, be connected to a continued preference of this verb for animate stimuli, but the attested instances with abstract or other forms of inanimate stimuli do not provide contrastive evidence of this, since they are not associated with personal constructions. Since the experiencer argument of *quemen* does not typically occur in preverbal position, its reanalysis as nominative subject was certainly not required for structural reasons, which may be another factor in the continuation of *quemen* as a transitive verb with a nominative stimulus argument. In the present corpus data, *quemen* is largely superseded by *plesen* during the ME period M2. As a consequence, it preserves the more literal meaning of pleasing throughout its attestation and does not exhibit any of the formulaic uses that were found with its near-synonym *liken*. This situation differs greatly from the one with *plesen*, which will be the subject of the following section.

4.2.2. Plesen

The earliest evidence of ME *plesen* ‘to please’ in the *PPCME2* comes from the *The Earliest Complete English Prose Psalter*, which is dated to around 1350 at the end of the period M2. Its attestation in the corpus is thus almost mutually exclusive with that of *quemen*, whose last instances in the *Ayenbite of Inwyrt* date to 1340. Etymologically, the verb goes back to OF *plaire* ‘to please’, although different stems are attested in Middle French and Anglo-Norman, both of

which are potential contributors to the ME loanword (cf. *OED s.v. please*). Since *plesen* enters the language around the same time that *quemem* becomes largely obsolete, it seems natural to regard it as the direct successor of the latter (cf. also Trips/Stein 2019: 244 *et passim*, with reference to Allen 1995). The general popularity of the verb, which is evidenced by its frequent attestation in the subsequent periods, entails a much wider range of syntactic and semantic contexts, however, not all of which have a direct model in the more restricted attestation of *quemem*. The most important among the assumed commonalities of the two verbs is the identification of their respective stimulus argument as a more volitional cause argument compared to a supposedly more neutral theme argument of *likem*. While this interpretation appears to be almost perfectly reflected in the high proportion of human stimuli of *quemem* and in its use in purpose clauses or as complement of verbs of volition and ability, the semantic and syntactic attributes of the complements of *plesen* are more varied and thus provide an interesting point of comparison with *likem*. Also, while *plesen* is not among the verbs whose experiencer argument underwent personalisation by the extension of nominative case but rather an example of the preservation of impersonal constructions by the introduction of an expletive into preverbal position, certain constructions of the verb are identified by Allen as favourable contexts for the early adoption of nominative case according to the semantic hypothesis of an agentive interpretation of the relevant participant. The investigation of these contexts will therefore be another point of discussion.

4.2.2.1. Database

The total data extracted for *plesen* are much more copious than those for *quemem* or, in fact, any of the other verbs that were investigated in this study. The number of observations derived from the *PPCME2* is comparable to that of *likem* with 168 compared to 188 instances, which already indicates a high incidence of *plesen*, since this verb is only attested from the end of M2 onwards. More importantly, though, the increased frequency in the *PCEEC*, which both verbs exhibit, is much more pronounced for *plesen* than it is for *likem*. The initial query of the *PCEEC* for the former verb yielded 2,520 results. After the exclusion of obvious false-positive hits, including *pleasure* and *place*, 2,350 potential instances of *plesen* remained. These were reduced to a random sample of 500 observations, all of which were found to be true instances of *plesen* during the process of annotation. While it is thus possible that the full set of 2,350 potential instances in the *PCEEC* contains forms that do not belong to *plesen* (MED33713), the chances of this appear to be low in view of the fact that all of the forms that were checked manually

were identified as tokens of the verb. An overview of the standardised frequency of *plesen* and *liken*, including the potential instances of *plesen* that were not checked, is given in figure 15 below, which also contains the frequency of *quemen* in M1 and M2 for comparison:

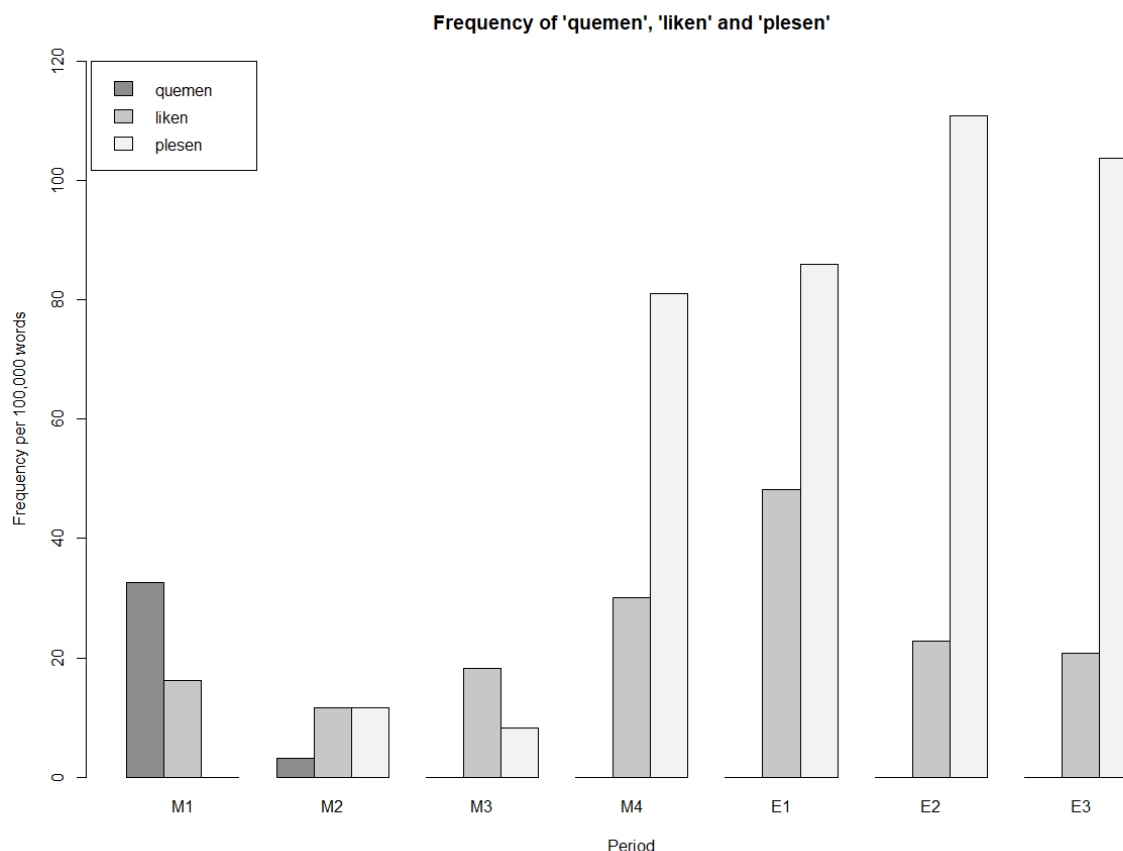


Figure 15 Standardised frequency of *quemen*, *liken* and *plesen* per 100,000 words.

The plot indicates a marked increase in the frequency of *plesen* in M4 compared to a moderate increase in the frequency of *liken* until E1. In the later periods, the frequency of *liken* declines again, while the frequency of *plesen* continues to rise and only declines moderately in the final period E3. A large share of this increase is attributable to the texts in the *PCEEC*, however. Table 13 below contains the standardised frequency of *liken*, *plesen* and *quemen* by corpus:

	<i>PPCME2</i>				<i>PCEEC</i>				
	M1	M2	M3	M4	M3	M4	E1	E2	E3
<i>liken</i>	16.3	11.7	16.9	16.7	46.1	44.7	48.2	22.7	20.7
<i>plesen</i>	-	11.7	8.2	30.9	10.3	136.1	86.0	110.9	103.7
<i>quemen</i>	32.5	3.2	-	-	-	-	-	-	-

As can be seen, the standardised frequency of *plesen* in M4 is more than four times higher in the *PCEEC* than it is in the *PPCME2* with about 136.1 compared to 30.9 instances per 100,000 words. Similarly, the frequency of *likem* in M4 is higher in the *PCEEC* with about 44.7 instances per 100,000 words compared to 16.7 instances in the *PPCME2*, but this is only about two and a half times as high and thus less pronounced than the difference between the two corpora for *plesen*. Interestingly, the frequency of *plesen* in M3 is lower than that of *likem* in both corpora, which indicates that the relative popularity of *plesen* had not been established by that period. Since the size of the relevant subsection of the *PCEEC* is particularly small, however, the difference between the corpora may not be robust. It should also be noted, once more, that the chronological distribution of the verbs is represented on the basis of the manuscript dates specified in the corpora rather than the dates of the original texts, which means that some of the instances may continue earlier uses of a given verb, provided they are not the result of lexical substitution. The absolute frequencies in table 14 below illustrate how the difference in the size of the subcorpora for each period is amplified by the relatively high standardised frequency of *plesen*. The data in E2, for example, which is the largest subsection of the *PCEEC* with 910,675 words, account for 1,010 instances of *plesen*, which amounts to about 43.0% of all observations from the *PCEEC* and about 40.1% of all observations from both corpora combined.

	<i>PPCME2</i>				<i>PCEEC</i>				
	M1	M2	M3	M4	M3	M4	E1	E2	E3
<i>likem</i>	42	11	68	67	9	163	149	207	115
<i>plesen</i>	-	11	33	124	2	496	266	1010	576
<i>quemem</i>	84	3	-	-	-	-	-	-	-
<i>plesen (sample)</i>					1	105	54	219	121

Table 14 also displays the distribution of the sampled observations that were drawn from the full set of potential instances of *plesen* in the *PCEEC*. Allowing for a certain amount of random variation, the distribution of observations in the sample is generally representative of those in the full set. The dataset of *plesen* that was obtained by sampling is somewhat smaller than that of *likem* with 668 observations compared to 831 observations. This process also reduced the dominance of the *PCEEC* in the final dataset from a ratio of about 14 : 1 to about 3 : 1, which makes the dataset more similar to *likem*, where the equivalent ratio is about 3.4 : 1.

As with *liken*, the observations of *plesen* belong to several different clause types. 63 or about 9.4% of the total 668 observations in the dataset of *plesen* were identified as non-finite clauses, which is roughly comparable to the total proportion of non-finite clauses of *liken* with about 6.9%. The composition of these instances is not the same for both verbs, however. Infinitival clauses are somewhat more frequent with *plesen* than with *liken* (c. 50.8% compared to 42.1% of all non-finite instances) and past participles are noticeably more frequent (c. 20.6% compared to 7.0% of non-finite instances). The higher frequency of the latter correlates with the proportion of passive constructions, which is much higher for *plesen* with about 21.0% of all observations (140 instances) compared to only 1.6% in the case of *liken*. The remaining instances of finite active clauses of *plesen* amount to 465 or about 69.6% of all observations, 3 of which do not contain an experiencer argument. Figure 16 below illustrates the diachronic distribution of the different clause types.

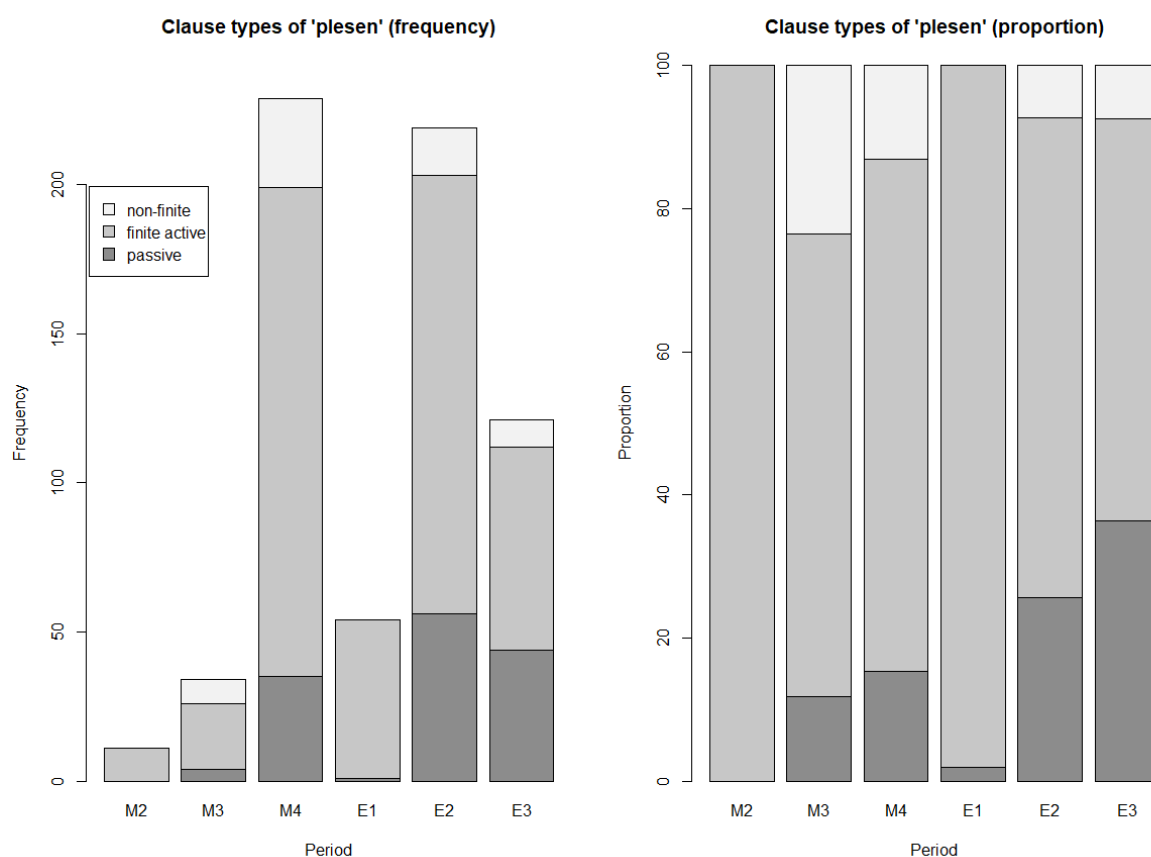


Figure 16 Clause type of *plesen* by period (n = 668).

Regarding the high number of passive constructions, it should be noted that the *PPCME2* and the *PCEEC* do not distinguish between participial and adjectival forms and that no such distinction was introduced in the present investigation either. Instead, the instances of *plesen* that were categorised as passives were mechanically identified based on the presence of a past-

participial form of the verb in conjunction with a form of *be*. Based on a cursory view of these instances, this combination can be tentatively interpreted as a combination of copular *be* and a derived adjectival form in the majority of cases. The most important reason for excluding passive constructions from the dataset of *liken* was the observable difference between active and passive clauses in terms of their complementation pattern, as one of the verbal arguments is typically suppressed in the passive construction. This is also the case with some of the passive/adjectival constructions of *plesen*, which did not occur with an expressed stimulus argument in 26 or 18.6% of all 140 instances. In about as many instances, however, the stimulus was explicitly realised by a PP in the template *pleased with X* (27 or 19.3% of all instances), which suggests that the syntactic construction under investigation primarily serves the modification of information structure rather than the suppression of one of the participants, since it allows the experiencer of *plesen* to occur as the utterance topic. Since this position is invariably held by the experiencer argument in the relevant constructions, the verbal semantics can be identified as causative, which is, of course, in keeping with the general interpretation of *plesen* in active clauses, but which contrasts with *liken*, for which a diachronic change from a causative to a receptive interpretation of the verbal semantics was observed. Finally, the most frequent type of complement of the passive/adjectival construction of *plesen* are propositional stimuli, which amount to 87 or 62.1% of all instances, including 1 instance deleted under coordination. Specifically, these take the form of infinitival clauses (79 observations) and of finite subordinate clauses (8 observations), the former of which are concentrated in the 17th century (E2–E3) and thus chronologically distinct from the other types of complements, which are distributed more widely between M3–E3.

4.2.2.2. Variables

4.2.2.2.1. Construction Type

The reduced dataset of *plesen* including finite active clauses with two arguments contains 462 observations. Of these, 328 (c. 71.0%) were identified as impersonal constructions and 27 (c. 5.8%) as personal constructions, while a relatively high number of 107 observations or about 23.2% remained ambiguous in terms of their construction type according to the criteria applied. As with *liken*, the most important criteria were case marking of the experiencer argument and constituent order in constructions with an expletive. Morphological ambiguity of the experiencer argument was a particularly frequent phenomenon in the correspondence corpus, where the experiencer, who frequently corresponds to the addressee, is often realised by the

originally oblique form *you* or one of its spelling variants. This form loses its distinctiveness over the course of its attestation, which is reflected in the fact that 84 or about 53.8% of the 156 instances of *you* encountered were classified non-distinctive based on an inspection of their respective context. In addition, NPs headed by full nouns were always morphologically ambiguous. The single most frequent item in this connection is represented by *God* and its lower-case spelling variant *god*, which exhibit a total number of 61 occurrences. Another important contributor to morphological ambiguity of the experiencer argument is the high frequency of instances of abstract reference. While individual terms such as *your Lordship* and *hir Majesty* are less frequent than the high-frequency experiencer *God*, the total number of items that were classified as abstract reference amounts to 100 or about 21.6% of the 462 observations, which is an even higher proportion than that observed in the equivalent dataset of *liken*, where 110 instances or about 17.3% of the total 745 observations were classified as such. These three types of experiencer arguments taken together already account for 245 observations in which the experiencer argument is morphologically ambiguous, but if all nominal experiencers and occasional instances of deleted constituents are included, this number rises to 300 or about 64.9% out of the total 462 observations, which implies a much greater potential for structural ambiguity than has been generally assumed for experiencer verbs. This relatively high potential for structural ambiguity is, of course, balanced by the frequent use of expletive *it*, which usually ensures postverbal position of the experiencer. 162 observations were thus classifiable as impersonal constructions based on the constituent order resulting from the use of an expletive, while only 150 observations were immediately classifiable as either impersonal or personal constructions based on case marking alone. An additional 12 instances were classified as impersonal constructions based on the realisation of the experiencer argument by a PP, while the more marginal criteria of the case-identity constraint and control of verb agreement accounted for 14 and 17 classifications respectively.

The construction of *plesen* according to the above criteria is predominantly impersonal, which is consistent with the traditional characterisation of this verb as one that favoured the introduction of an expletive constituent over the adoption of nominative case for the experiencer argument. The limited number of personal instances that occur are mainly found in the later EModE periods E2 and E3. The complete diachronic distribution of the construction types is illustrated in figure 17 below, which also includes ambiguous observations, since these facilitate a better appreciation of the data. The proportion of impersonal and personal constructions in E3, for example, is virtually even, which suggests a significant increase in personal constructions compared to earlier periods when considered in isolation. This proportion is based

on a very low number of classifiable observations, however, since only 13 observations were identified as impersonal constructions and 12 as personal constructions, while the greatest share of observations in that period is, in fact, ambiguous. The reason for this is a significant decline in the use of the expletive, which thus no longer disambiguates the high proportion of morphologically ambiguous experiencer arguments. Similarly, the notably high proportion of ambiguous observations in M3 can be attributed to the virtual absence of expletive constituents, which do not become prominent with *plesen* until the subsequent period M4. This contrasts with *liken*, for which instances with an expletive were more numerous already during the period M3, and it illustrates once more the importance of the expletive for the morphosyntactic disambiguation of the construction of *plesen*:

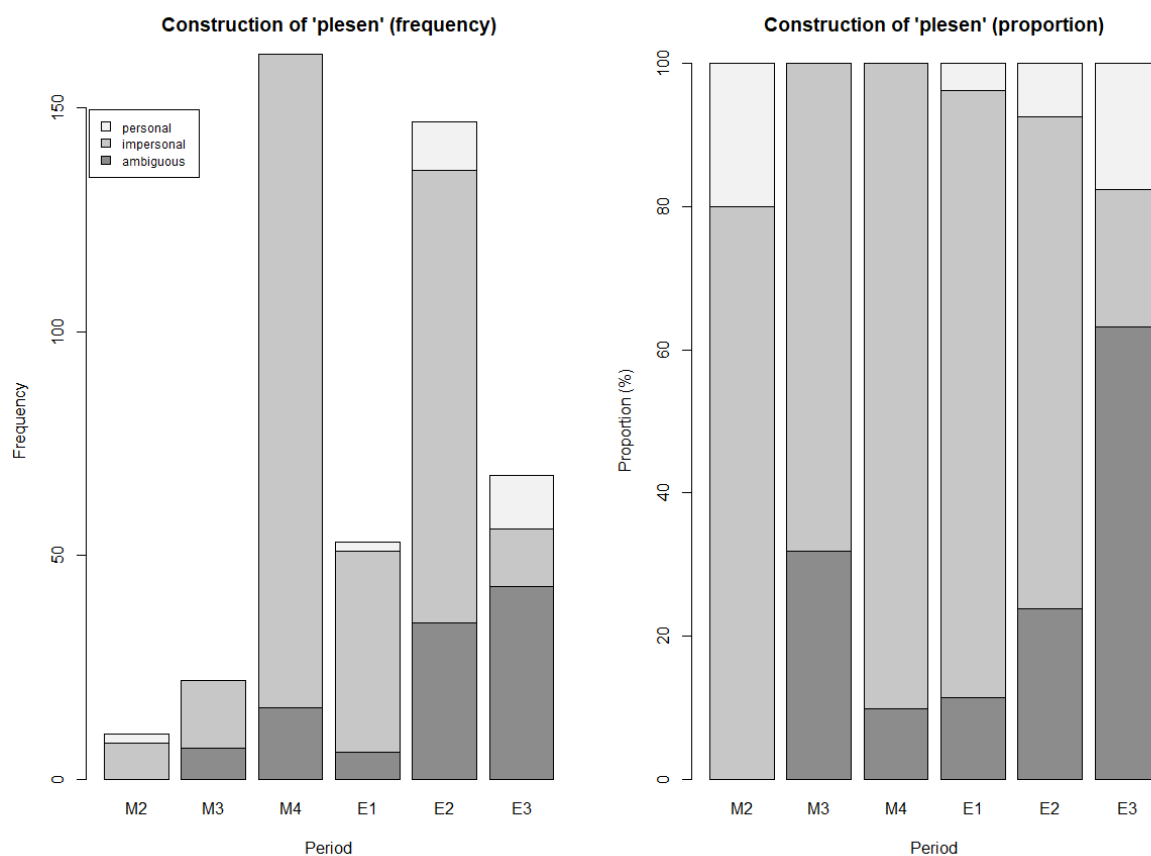


Figure 17 Frequency and proportion of the construction of *plesen* ($n = 462$).

Regarding the observations that were identified as personal constructions, the occurrence of 2 early instances in M2, which account for 20% of the 10 observations in that period, is notable. Both instances derive from *The Earliest Complete English Prose Psalter*, and they are thus potentially suspect of being influenced by the Latin exemplar underlying their translation. Nevertheless, they are at least consistent, since they both exhibit the use of a nominative experiencer in combination with a PP stimulus headed by the preposition *in*, cf. e.g. *and ich*

plested in þy sobenesse ‘and I delighted in thy truth’ (CMEARLPS,28.1180). This use of a PP stimulus with *pleesen* in a finite active clause equally sets the two observations apart from the remainder of the dataset and suggests that they are distinct from the later occurrence of nominative experiencers with *pleesen* evidenced from E1 onwards. Whether or not they are to be regarded as genuine instances of vernacular Early Middle English is difficult to judge, but since their occurrence falls into the earliest period of the adoption of *pleesen*, the syntactic adaptation of the loanword may have been not yet entirely fixed. This is also suggested by the remaining instances in the *Prose Psalter*, which are mainly constructed with a PP experiencer and thus equally distinct from most of the later observations of *pleesen*. The personal observations from the EModE periods E1–E3, on the other hand, are exclusively constructed with propositional stimuli, most of which are inferred from the context, cf. e.g. [...] *any woman that had witt and discretion might make an Asse of him and Govern him as shee pleased*, (OSBORNE,110.047.2577). The inferred stimulus argument in this instance can be construed as an infinitival clause [*to govern him*], which is also the most typical complement of personal constructions in the later periods, while much fewer instances were construed as finite-clauses with a non-identical subject. The majority of personal instances are thus found in contexts that correspond to Allen’s NO PROP construction or, in the present terminology, in constructions with an inferred propositional stimulus argument.

4.2.2.2.2. Semantic and Syntactic Properties

The typical degree of animacy of the stimulus argument of *pleesen* changes diachronically as it does with *liken*, but the emerging trend is the opposite of that observed for the latter verb. Figure 18 below presents the frequency and proportion of human and non-human stimuli in the six periods in which *pleesen* is attested, beginning with M2. Since the purpose of this information is to characterise the diachronic change in the properties of the verbal arguments of *pleesen* rather than to illustrate the database of a regression model, there is no need to restrict the observations to unambiguous finite active clauses. Instead, the figure represents the full set of *pleesen*, including passive/adjectival constructions and non-finite instances of the verb, the majority of which are constructed with two explicit participants and thus allow for an interpretation of the animacy of both arguments. For the sake of comparability, the figures of the reduced set of finite active clauses with two arguments are cited alongside those of the full set in table 15 further below. It should be noted that the full set includes 10 observations with a missing experiencer argument, 3 of which take the form of finite active clauses. These were removed

from the reduced set in addition to the non-finite and passive instances. 36 observations with a missing stimulus argument, on the other hand, were removed also from the full set, since they do not provide any information about the properties of the stimulus argument. The data from the full set thus comprise 632 observations and the reduced set including ambiguous constructions 462 observations:

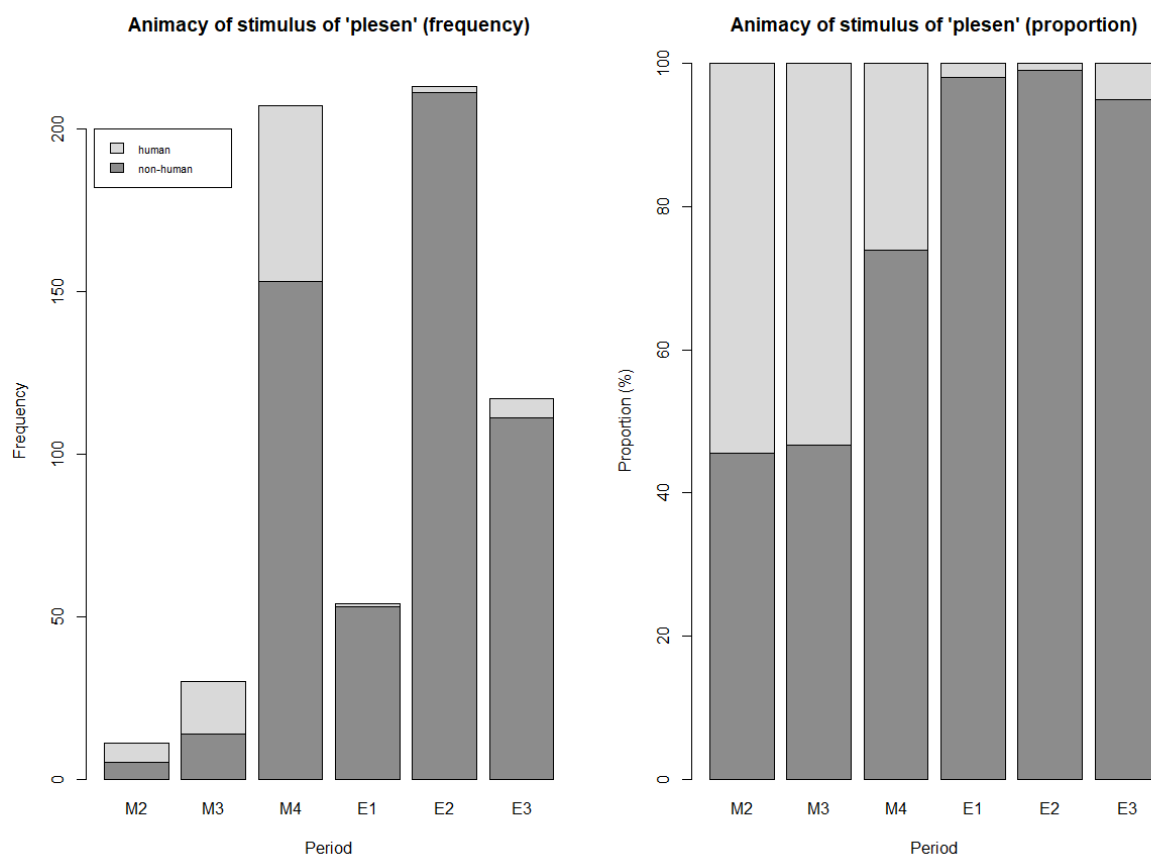


Figure 18 Animacy of the stimulus argument of *plesen* ($n = 632$).

The figure illustrates a relatively clear divide between the ME periods M2–M4 and the EMode periods E1–E3. During the former, the proportion of human stimuli shows a decline from about 54.5% in M2 and 53.3% in M3 to about 26.1% in M4, but the proportion is markedly lower in the later periods, where it ranges between 0.9% in E1 and 5.1% in E3. This diachronic divide is notably correlated with the two corpora, however, which is apparent from the fact that 52 instances or about 96.3% of the 54 human stimuli in M4, which is the only period that contains a substantial number of words from both corpora, derive from the *PPCME2* and only 2 instances or 3.7% derive from the *PCEEC*. The absolute frequencies that correspond to the proportion of human and non-human stimuli in the different periods are given in table 15, once for the full set underlying the two plots in figure 18 above and once for the reduced set of finite active clauses with two arguments. The inclusion of non-finite instances and passive/adjectival

constructions in the full set affects the proportion of human and non-human stimuli in several different ways. The most notable difference is the much higher frequency of non-human stimuli in E2 and E3, which results from the association of passive/adjectival constructions in these periods with infinitival stimuli. Since a small number of additional instances of human stimuli are found in non-finite clauses in these periods as well, however, the proportion of human stimuli is, in fact, slightly higher in the full set compared to the reduced set, which contains only a single observation with a human stimulus in each of the periods E1–E3:

		M2	M3	M4	E1	E2	E3
full set (n = 632)	non-hum	5	14	153	53	211	111
	human	6	16	54	1	2	6
reduced (n = 462)	non-hum	5	10	131	52	146	67
	human	5	12	31	1	1	1

In the periods M2 and M4, the proportion of human stimuli is equally higher in the full set compared to the reduced set, which is due to the fact that the null constituents in infinitival clauses excluded from the latter set are typically coreferential with a preceding human participant. In M3, on the other hand, the removal of non-finite instances leads to a slight increase in the proportion of human stimuli, both when compared against the full set and against the preceding period M2 because M3 contains a somewhat higher number of present participles than infinitival clauses. These details aside, the most remarkable fact is the sharp decrease in the proportion of human stimuli in the *PCEEC*, which is apparent in both the full and the reduced set and only balanced somewhat in the former by the presence of non-finite instances with human stimuli in the final period E3.

Apart from the distribution of human stimuli, which constitute one of the four semantic groups investigated besides non-human concrete, abstract, and propositional stimuli, the overwhelming dominance of propositional stimuli of *plesen* is striking. Within the full set of *plesen* (n = 632), 472 or about 74.7% of all observations involve a semantically propositional stimulus argument, which is typically realised syntactically by an explicit or inferred non-finite clause or by an anaphoric pronoun. This proportion is even higher than the equivalent proportion of propositional stimuli with *liken*, which amounts to about 51.5% in the set including non-finite instances and passive constructions (n = 831). Abstract and non-human

concrete stimuli, on the other hand, are relatively marginal with *plesen*, constituting 39 (c. 6.2%) and 36 instances (c. 5.7%) respectively. The diachronic distribution of the semantic types is given in figure 19 below.

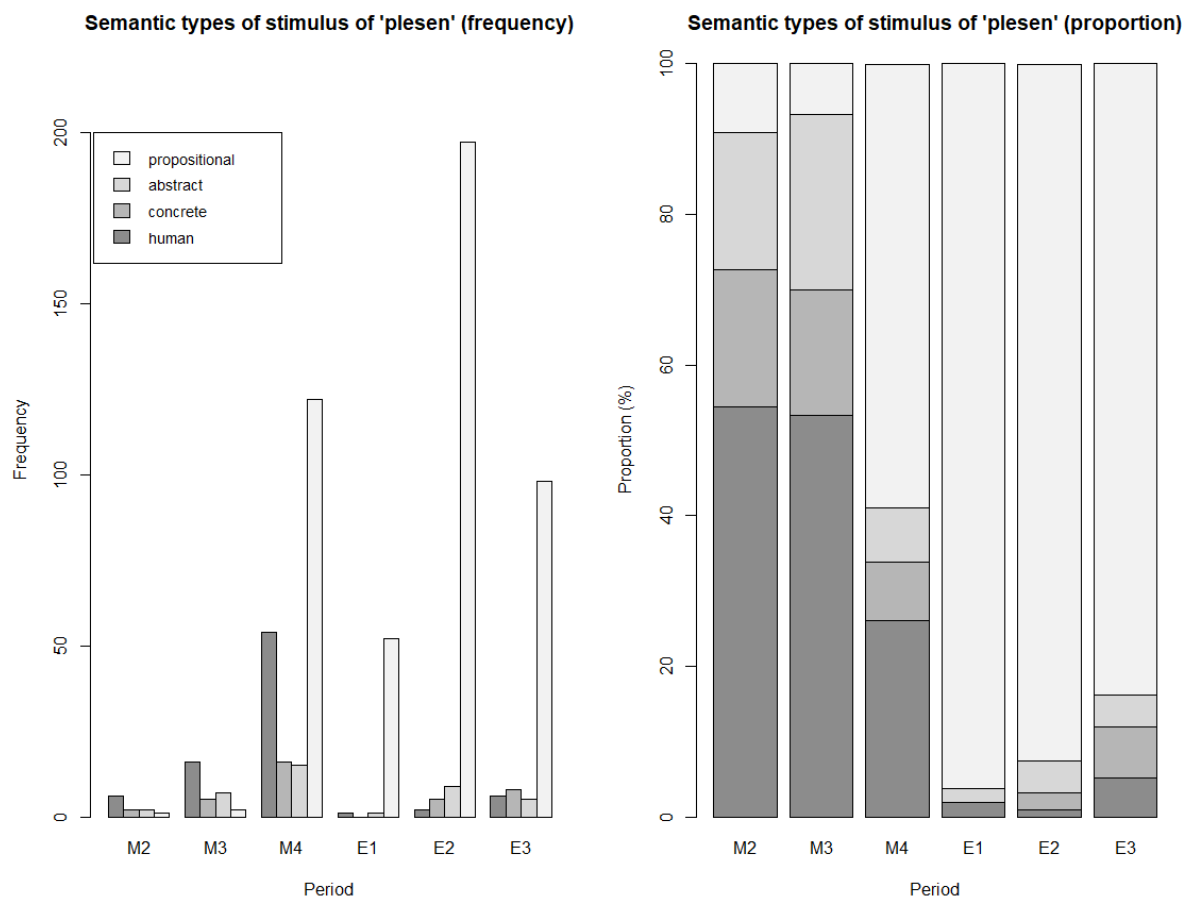


Figure 19 Semantic types of the stimulus argument of *plesen* (n = 632).

The relative increase in the number of propositional stimuli from M4 onwards is particularly marked compared to the increase of all non-human stimuli combined, since only 1 (c. 9.1%) and 2 (c. 6.7%) instances of propositional stimuli are found in the earlier periods M2 and M3 respectively. The total number of 11 observations in the earliest period M2 is, of course, very low, which means that the proportions obtained for that period are not necessarily reliable. Nevertheless, they match the equivalent proportions in M3, which is represented somewhat more favourably with a total of 30 observations containing a stimulus argument. The highest proportion of propositional stimuli is found in E1 with 52 observations (c. 96.3%). This period is notable for a near absence of all other types of stimuli and contains only 1 human and 1 abstract stimulus argument (c. 1.9% each) besides the propositions. Towards the end of the investigated period, the proportion of non-propositional stimuli increases again, although this increase is slight, with 6 human (c. 5.1%), 8 concrete (c. 6.8%) and 5 abstract stimuli (c. 4.3%) compared to 98 propositions (c. 83.8%).

The categorial realisation of the stimulus argument of *plesen* largely corresponds to the facts that have already been mentioned and hence does need to be discussed in detail. The inclusion of passive/adjectival constructions and past-participial clauses in the full set underlying figure 20 is reflected in the occasional realisation of the stimulus argument by a PP, which is uncommon in finite active clauses with the exception of two isolated instances in M2. This situation contrasts with that of *liken*, for which the realisation of the stimulus by a PP constituent was a relatively frequent phenomenon also in finite active clauses in E2:

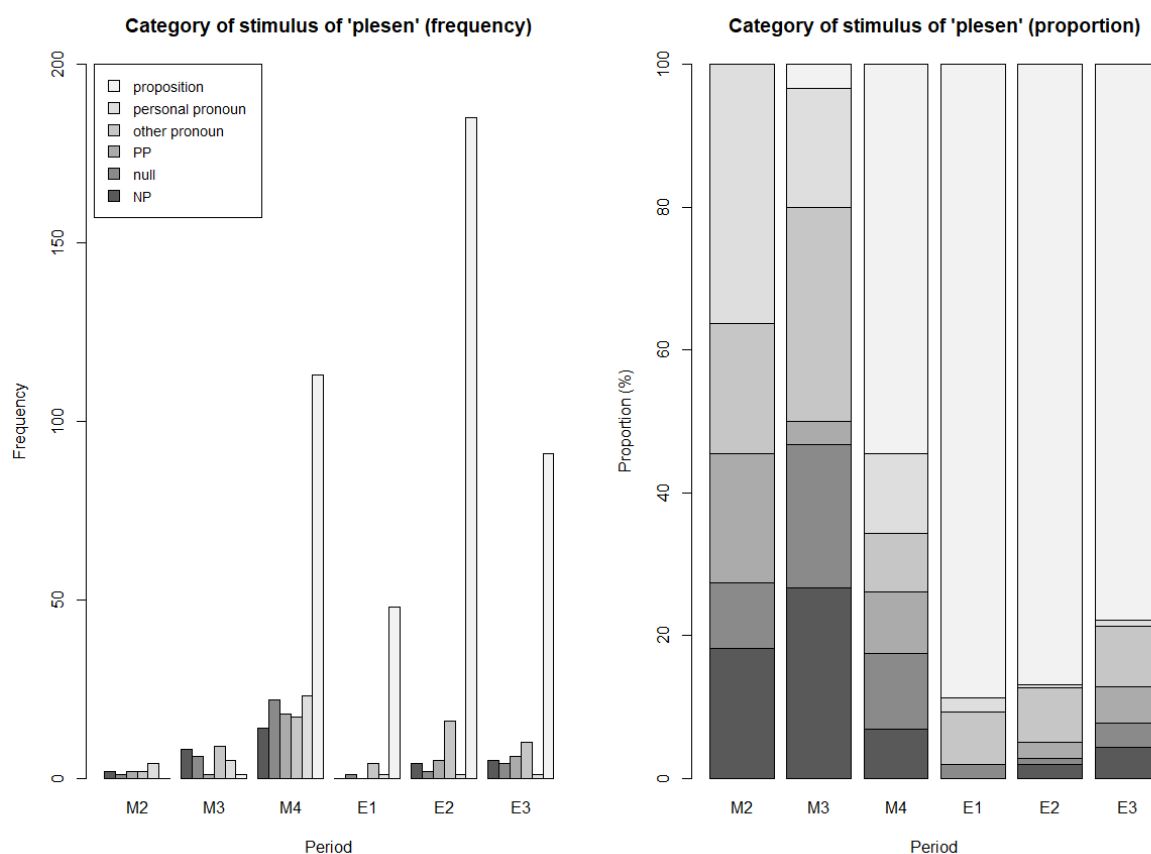


Figure 20 Category of the stimulus argument of *plesen* (n = 632).

Another fact that emerges from figure 20 is the lower proportion of propositional stimuli in terms of categorial realisation when compared to that of semantically propositional stimuli given in figure 19 above. The former ranges between 88.9% and 77.8% in E1–E3, while the latter ranges between 96.3% and 83.8% in the same periods. The difference is largely due to what can be called prefacing or appositive uses of *plesen*, which are not integrated syntactically into the surrounding context, but for which cataphoric reference of the pronominal stimulus argument to the subsequent proposition can be assumed. This analysis creates an inconsistency with the analysis of similar instances of *liken*, which were interpreted as inferred stimuli in conjunction with an expletive (cf. section 3.3.3.2). The two opposing analyses are contrasted

further below. Since the relevant instances of *liken* were much less numerous, however, the difference in analysis should not impact the overall findings too greatly. An example of the construction with *plesen* is given in (153):

(153) May it please **your most Excellent Majesty** I presume to send herewith a copy of my humble petition to the Kinges Majesty wheareby your Majesty may perceave with lesse trouble then any other relation of mine as much in effect as I can say of the condition of my present estate and harde fortune . (STUART,248.059.1063)

While the present instance could also be construed as an embedded complement clause with a zero complementiser, similar instances for which such an interpretation is less likely suggest that the clause containing *plesen* is, in fact, a kind of apposition to the information that follows. Most typically, these instances occur at the very beginning of a given letter or at the beginning of a new section. They thus appear to serve the combined function of added politeness and the organisation of the discourse structure. Their conventional nature is apparent from the relative formalism of their lexical and syntactic components, which is comparable to similar formulaic constructions of *liken*. The syntactic analysis applied to the appositive instances of *plesen* differs from superficially similar instances in which the stimulus argument is realised by a non-finite or finite subordinate clause preceded by an expletive pronoun. In the appositive instances, this pronoun was, instead, interpreted as a cataphoric item that refers to the propositional content of the subsequent but syntactically separate clause. This analysis is certainly not ideal for all cases, some of which are transitional between appositive clauses like the one in (153) and the frequent formulaic expressions in the correspondence corpus involving verbs like *witen* ‘to have knowledge’ or *understonden* ‘to learn’ that were already encountered with *liken*. For an example of these instances cf. (154) below:

(154) Fordermore , Syr , yeff yt plesse **your maystershipe** for to understonde þis , I have ressayvid your wollys as ffayer and as hole as any mannys in the fflete .

‘Furthermore, Sir, if it please your mastership to understand the following – I have received your woolls as fair and as whole as any man’s in the fleet’
(STONOR,II,4.041.627)

Apart from the presence of an infinitival stimulus *for to understonde þis*, this instance is equivalent to the one in (153), but since it belongs to the same clause as *plesen*, this changes the interpretation of pronominal *yt* from a cataphoric pronoun to an expletive. With regard to the semantic interpretation of the stimulus argument, this does not change much, of course,

since the stimulus is analysed as propositional in either case, but, in terms of its categorial realisation, the analysis results in the aforementioned difference between clausal constituents and anaphoric pronouns, which impacts the observed frequency of these categories. Also, since the analysis of the construction type according to the criterion of constituent order was restricted to instances with an expletive, this criterion was not applied to the appositive instances, despite their superficial similarity with impersonal constructions. There are, thus, a certain number of observations of *plesen* that could increase the proportion of impersonal constructions in the later periods even further if their analysis was adapted. The appositive clauses and the related formulaic uses of *plesen* will be discussed further in the following section, which will not only provide a summary and interpretation of the diachronic development of this verb but also compare its development to the near-synonymous verbs *quemen* and *liken*.

4.2.2.2.3. Discussion

Even though no general change from impersonal to personal constructions or from causative to receptive verbal semantics is observable with *plesen*, there are several points to note about the diachronic development of this verb. Its first traces in *The Earliest Complete English Prose Psalter* and its later attestations in M3 suggest that the newly acquired loan verb corresponds to the semantic scope of both the indigenous verb *liken* and the recessive verb *quemen*, as far as this can be judged by the typical semantic properties of their respective stimulus arguments. On the one hand, *plesen* continues the construction with more or less volitional human stimuli that is typical of *quemen*, although the narrow sense of pleasing God by behaving in accordance with Christian doctrine is not as dominant as it is with the latter verb. Instead, *plesen* also continues the use of concrete and abstract stimuli that is more commonly found with *liken*. The characterisation of *quemen* and *liken* according to the typical semantic properties of their stimulus argument is not entirely straightforward, however, and requires some elaboration. It is certainly true that *liken* exhibits a low proportion or even near-absence of human stimuli in the earliest periods. The few instances of human stimuli that do occur appear to be mostly genuine and comparable to similar instances with *quemen* and *plesen*, but they are still much less typical of this verb than they are of *quemen*. In the *Trinity Homilies*, which account for most of the relevant instances, non-human stimuli occur in 9 or about 64.3% of all 14 instances in the full set of *liken* including ambiguous constructions and non-finite instances, while *quemen* is consistently constructed with human stimuli in all 9 instances from the same text. Of course, the low number of attestations and the dominance of religious literature in the early subsections

of the *PPCME2* make it difficult to judge the significance of these observations for the general use of *liken* and *quemen* at the time. Since a major function of the homiletic literature is to provide guidance about God-fearing behaviour, the frequent combination of *God* as the experiencer argument and human referents as the stimulus argument of *quemen* may also be an epiphenomenon of the prominent subject matter. In the *Ormulum*, which contains most of the observations of *quemen*, the combination is particularly frequent in the almost formulaic expression *[p]a menn þatt wel himm cwemenn* ‘the people that please him [sc. God] well’ (e.g. CMORM,I,190.1563), which is repeated several times in slight variations. The stronger association of *quemen* with human stimuli, which Allen has attributed to the lexical specification of a more agentive or volitional cause argument and which is suggested contextually by additional linguistic facts such as the verb’s frequent conjunction with instrumental adjuncts or purpose clauses, thus seems to be connected to the presence of a specific subsense of gratification or compliance, which is suitable in the context of religious laws and which appears to have been present for *quemen* but less so for *liken*. Situations involving actual emotional fondness of the experiencer for another human participant appear to be virtually absent in the early sections of the *PPCME2* or at least not covered by the verbs under investigation. It is thus difficult to demonstrate whether *quemen* or *liken* would have been the more appropriate verb to occur with human stimuli in such contexts. It is clear, however, that the animacy of the stimulus argument in such cases does not necessarily entail the same agentive interpretation that is inherent in the gratification described by *quemen*, and it thus seems probable that *liken* would thus be eligible to occur in the relevant contexts if these were represented in the early sections of the corpus. Later occurrences of *liken* illustrate this usage, although they can no longer contrast with *quemen*, which is obsolete at this stage, cf. e.g. the following instance from M4: *[t]he kyng lyked and loved this lady wel*, (CMMALORY,2.12). Yet even if the near-absence of human stimuli with *liken* in the early ME periods is regarded as the result of an absence of suitable contexts in the corpus rather than as a lexical restriction, the occurrence of *plesen* with both human and non-human stimuli in the available contexts essentially covers the combined distribution of *quemen* and *liken*.

There is one exception in the distribution of *plesen* compared to *quemen* and *liken*, however, and that is the occurrence with propositional stimuli, which is attested for *liken* but not typical of *plesen* in the earlier periods. Even if semantically propositional stimuli are included in the figures, there is still a marked difference between *plesen* with only 3 instances in M2 and M3 combined and *liken* with 50 instances in M3 alone when all available observations are considered. The early increase of propositional stimuli with *liken* in M3 is

particularly notable compared against their consistently low proportion with *plesen*. It is largely accounted for by a range of clause types that convey some aspect of indefiniteness, especially conditional clauses (*if you like*) and comparative clauses (*as you like*), but also temporal, locative or other clauses with an indefinite pronoun (*when / where / what(ever) you like*). With *plesen*, such instances do not occur until the subsequent period M4, which, in addition, exhibits a clear divide between the *PPCME2* and the *PCEEC*. The former corpus contains most of the human stimuli recorded in that period, while the *PCEEC* contains mostly propositional stimuli and only 2 instances of human stimuli. These propositional instances also include cases of the indefinite type with an inferred stimulus argument, typically realised in conjunction with an expletive (*as it pleased him*), but, more frequently, a specific stimulus argument is expressed, which generally refers to a request directed at the addressee or provides some kind of formulaic introduction to new information as a means of structuring the discourse. Propositional stimuli remain the most prominent type with *plesen* throughout the remainder of the investigated period, while equivalent instances with *liken* decrease notably in E2 and are almost absent in E3. The developments of the proportion of semantically propositional stimuli with *liken* and *plesen* are compared in figure 21:

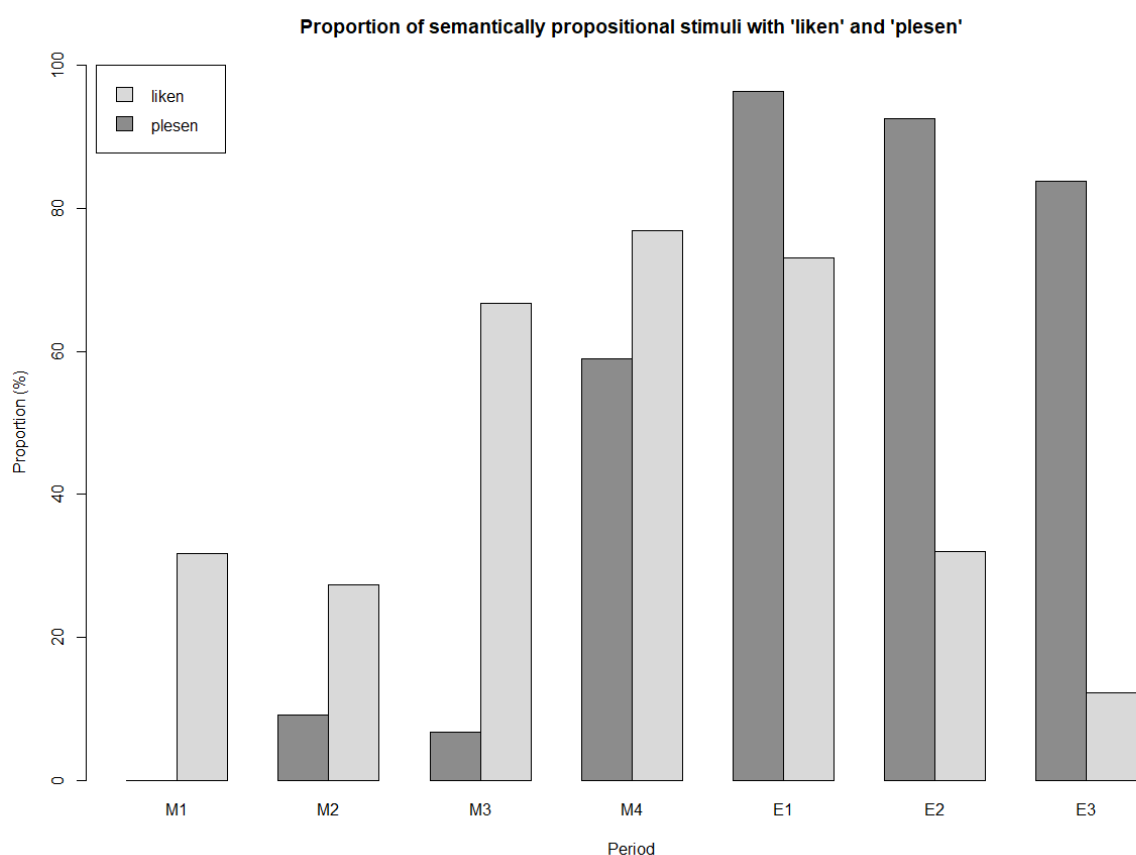


Figure 21 Semantically propositional stimuli with *liken* (n = 822) and *plesen* (n = 632).

The divide between the two corpora regarding the distribution of the different types of stimuli in M4 illustrates clearly that the increase of propositional stimuli in that period is correlated with the shift in the dominant genre from religious literature to correspondence. If only the data from the *PCEEC* are considered, the proportion of propositional stimuli with *plesen* would be almost as high (86.7 %) as that with *liken* (88.3%). The fact that the difference in the combined corpora is more pronounced is thus mainly due to the fact that non-propositional stimuli in M4 remain more frequent in the *PPCME2* with *plesen* (66.1%) than with *liken* (51.5%). The equivalent figures in M3 nevertheless suggest that *liken* leads the development of an increased use of propositional stimuli in both corpora, since 9 instances or 18.0% of the 50 propositional stimuli with *liken* in that periods derive from the *PCEEC*, while the same corpus does not contain a single observation of *plesen* from that period.

Similar to the increase with *liken*, the increased use of propositional stimuli with *plesen* in the *PCEEC* reflects the heightened degree of politeness that is characteristic of the semi-formal correspondence in that corpus. Accordingly, a shift in the main function of the verb from expressing the literal act of pleasing or gratification to a semantically bleached marker of politeness can be observed. This is particularly apparent in hortative contexts, which are indicated by modal items such as *may*, *will* and *would* or, at times, by the imperative, cf. e.g. *[b]e pleased to present my h. [i.e. humble] service to the Reverend Mr. Wrench .* (BASIRE,256.012.491) and *[...] i desire you would bee pleased to let mee Read y=or= [i.e. your] iudgment on his letter* (FERRAR,308.036.755). In such contexts, the realisation of the proposition indicated by the stimulus argument is hypothetical, and its realisation depends on the performance of the experiencer, who is expressed by the coreferential PRO subject of the infinitival clause. They are thus typical contexts in which politeness towards the addressee or a third-person participant presents a requirement. Comparable constructions of *plesen* are also used in factual contexts of past events, however, whose outcome is no longer under the influence of the experiencer, cf. e.g. *[a]nd , in a speech to y=e= L=ds= [i.e. the Lords], he wase pleased to take notice y=t= [i.e. that] he had been misrepresented [...]* (HATTON,I,151.048.1182) or *[...] when I last attended your Ladyship yow were pleased to dyrect mee to enforme yow how my sister Saule made her demaund of one hundreth pounds [...]* (WENTWOR,218.068.944). The former of these citations is particularly interesting, since the addressee's observation about being misrepresented is clearly incompatible with the literal notion of pleasure that *plesen* would otherwise be expected to imply. Instead, the use of the verb appears to serve as an acknowledgement of the fact that the realisation of the action expressed in the subordinate clause was subject to the experiencer's choice. This kind of

acknowledgement is also present in the second citation, in which the addressee's instructions to be informed by the author are probably not the source of actual pleasure but merely characterised as being subject to the experiencer's discretion. Both instances are thus polite references to the influential position of the addressee and employed in recognition of their elevated status.

The relative semantic bleaching of *plesen* as a marker of politeness is also apparent from the frequent formulaic use of the verb, which is equally reminiscent of comparable uses of *liken*. The two formulaic expressions that were recognised for the latter verb as criteria for the identification of the construction type in section 3.4.1.2.2 occur in largely identical form with *plesen*. The formula *like / please (it) E (to) wit / know (that)*, which is chronologically prior and occurs predominantly in M4 with both *liken* and *plesen*, is associated with the function of introducing a new discourse topic. This is not only apparent from the realisation of the infinitival complement by verbs such as *witen* and *understonden*, which denote the acquisition of knowledge, but also from the frequent conjunction of this expression with sentence adverbs such as *item* 'also' and *furthermore*, which signal a change in topic. The second expression, *it may like / please E + inf.*, in which the modal verb *may* replaces the subjunctive, is mostly attested in E1 and E2 and is less restricted in its function. Its frequency could thus also simply be regarded as the result of creative language use, although the occurrence of instances with a similarly discourse-organising function suggests that the expression had a comparable formulaic function in the correspondence of the time, cf. e.g. *[i]t maie please your Lordship to be advertised that ther have bene with in this 8 or daies two severall shippes [...]* (PAGET,116.026.651) or *yt may also please your most Royall Mageste to knowe [...]* (CROMWEL,I,373.020.232). The most progressive form of this formula appears to be the appositive use discussed above, in which the expression *it may please E*, whose experiencer argument is realised either by pronominal *you* or various abstract terms of references, precedes the subsequent statement without explicit syntactic integration. These are typically found in E2 and E3 and thus appear to be a later development that emerges as a result of the increasingly conventionalised use of the second formula. A number of transitional instances indicate that the appositive use can be regarded as an abbreviation of the fuller use with an infinitival complement, cf. e.g. the instance in (155):

- (155) [...] After all dew recomendacions , plesethe your lordschip to be aduerticyd . Os I cam in Yorkschir tawardis Berwike the most party of gentilmen of the same , os Sir John Hastingis &c , held me cumpany , [...] (FOX,20.002.16)

The fact that infinitival complements such as *to be aduerticyd* can be omitted from the appositive use is a sign of its conventionalisation, which is also indicated by the frequency of the occurrence of both the appositive use and the fuller formulaic expression. It is difficult to distinguish between the function of *liken* and *plesen* in these formulaic expressions, since the two verbs appear to be virtually interchangeable. An investigation into their distribution with individual authors or regions may yield additional insights, since the correspondence by Thomas More, for example, exhibits frequent formulaic instances of *liken* but very few uses of *plesen*, while *The Cely Letters* exhibit numerous formulaic instances of *plesen* but none of *liken*. Overall, however, it seems that *plesen* enjoyed greater popularity than *liken* in constructions with a propositional stimulus argument, since its frequency in absolute terms is considerably higher and the present discussion is only based on the sample that was investigated in detail.

The point in which the development of *plesen* diverges most clearly from that of *liken* is the very limited development of personal constructions. These are entirely restricted to instances with a propositional stimulus argument and continue the indefinite type with an inferred stimulus (Allen's NO PROP construction) rather than the specific requests or polite formulae discussed in the preceding paragraphs. In the personal constructions, this type contains a nominative pronoun instead of the expletive, cf. e.g. *I doe expect my brother Robert and my sister , who may come in the coach with Jugg if they so plesse*. (BARRING,240.182.3164). Apart from 2 earlier observations in E1, the personal constructions are concentrated in E2–E3. Since instances containing the second-person plural pronoun *you* are typically morphologically ambiguous at this stage, these were analysed on the basis of verb agreement, which presupposes that the relevant verbal forms are not in the subjunctive mood. Personal constructions of this type also occur with *liken*, but these are much more sporadically distributed between the periods M4–E3 and never constitute the dominant type. The indefinite type is thus not generally at the vanguard of personal constructions across both verbs, even though the experiencer's coreferentiality of the PRO subject of an inferred infinitival clause containing an action verb may, indeed, be construed as an indirect implication of a higher degree of volition. The experience of emotion elicited by the performance of such an action is not a volitional act in itself, however, and the interpretation of the experiencer as more volitional in such cases appears to be connected to a particular subsense, which comes closer to that of taking a decision than to experiencing an emotion. Since this subsense can also be identified in constructions of *plesen* with an explicit propositional stimulus argument, the indefinite type is not a requirement for the suggested semantic interpretation. Instead, the conspicuous restriction of personal constructions of *plesen* appears to be the result of different strategies of realising the preverbal

position of a nominative experiencer argument. During the course of the later EModE period, both verbs exhibit a marked decline in the use of the expletive, which is completed somewhat earlier for *liken* than for *plesen*. While *liken* adopts the recently acquired personal construction with a nominative experiencer argument also for propositional stimuli, *plesen* exhibits a marked increase in passive/adjectival constructions instead, which equally result in the preverbal position of the experiencer argument. Since passive constructions appear to be avoided with inferred stimuli (**as he was pleased*), this would explain the conspicuous restriction of personal constructions of *plesen* to these contexts. Overall, the different developments of *plesen* and *liken* thus appear to align with differently lexicalised perspectives of the two verbs. With *liken*, the conceptual focus lies on the experiencer, whose topicality is reflected in its frequent preverbal position and the relative variability of the semantic properties of its corresponding stimulus argument. The conceptual focus of *plesen*, on the other hand, lies on the stimulus argument, which is more frequently preverbal and also more frequently animate. This difference is neutralised in constructions with a propositional stimulus argument, in which both *liken* and *plesen* are used as semantically bleached markers of politeness. Alongside the decline of constructions with an expletive constituent, this results in the extension of personal constructions to utterances involving propositional stimuli with *liken* and the increased use of passive/adjectival constructions with *plesen*.

4.2.3. Reuen

The inclusion of ME *reuen* ‘to regret, feel pity’ in the present investigation provides two potential points of comparison with *liken*. On the one hand, the emotion designated by this verb can be conceptualised as a more active process of cognition, which primarily applies to *reuen* in the sense of feeling regret or remorse. A strong emotional impact on the experiencer would probably not be conceived of as volitionally controlled, but active repentance as the result of a contemplative process may well result in greater chances of the experiencer to be identified as agentive. On the other hand, *reuen* in the sense of feeling pity or compassion for another participant is typically applicable to animate participants only, regardless of whether these take an active part in causing the emotion. This means that the hypothesised inference of agency from the semantic property of animacy may not be transferrable to all instances of *reuen* in equal measure. In addition, the prevalent identification of the verb with either one of these two subsenses will depend, to some extent, on the text types and the dominant topics represented in the two corpora. Both senses appear to be generally suitable for the religious literature of the

PPCME2, while their distribution in the *PCCEC* would seem to be contingent upon the specific topics discussed in the correspondence. This difference is, in some sense, reflected in the respective attestation of *reuen* in the two corpora, which will be presented in following paragraphs.

4.2.3.1. Database

The total data extracted for *reuen* are extremely limited, even when compared against the already limited data for the recessive verb *quemmen*. While the initial query returned a sizeable number of hits from each corpus, the majority of these were not tokens of the desired verb. The low precision of the character string used for the query was partly due to the double nature of the root-final grapheme <u>, which yielded a large number of false-positive hits in *re-v*^o such as *revenge*, *revoke* or *reveal* besides a number of other cases. After their removal, only 46 observations remained in the dataset, 41 of which were identified as forms of *reuen* and 5 as forms of *reusen* ‘to repent, pity’ with the alternative stem *reus-*. Since the two verbs are near-synonymous, and since their low frequency does not warrant an individual treatment, the observations of *reusen* will be discussed alongside those of *reuen*. There is an obvious difference, however, in that the forms of the former verb, which are restricted to the earliest period M1, comprise exclusively non-finite forms, while the forms of the latter comprise only finite forms in active clauses, which equally occur predominantly in M1 (31 or instances), although a small number of instances are also recorded in the later periods M3 (3 instances), M4 (6 instances) and E2 (1 instance). The verb is almost absent from the correspondence of the *PCCEC*, which accounts for only 2 of the later instances, while the 39 remaining instances of *reuen* derive from the *PPCME2*. The attestation of this verb thus displays a clear propensity towards the early religious literature, with the *Ormulum*, the *Trinity Homilies* and *Vices and Virtues* accounting for 28 of the 31 observations in M1. It is also worth noting that 7 of the 12 observations in the *Ormulum* are part of an enumerative section of the poem, in which *Himm reoweþþ* is repeated line-initially as a kind of stylistic device (cf. CMORM,I,192.1577–1583). The representativeness of the observations of *reuen* is thus limited further by the relatively small number of texts from which they derive and the stylistic repetition in at least one of these texts. In addition, the concentration of observations in the earliest period M1 makes it difficult to infer any diachronic tendency from the dataset. Nevertheless, the construction type of the observations of *reuen* and the properties of its arguments will be explored in the following sections. Apart from the non-finite instances of *reusen*, this survey excludes 1 observation with

a missing experiencer argument and 1 observation with a missing stimulus argument. The former is a clear instance of a general experiencer (cf. *for oftyn tyme rape rueth*, ‘because haste often causes regret’ (PASTON,I,337.110.3396)), while the latter contains a potential PP stimulus argument *after hire daedes* ‘after her deeds’ (cf. CMVICES1,103.1241), which was interpreted more conservatively, however, as a temporal adjunct. After the exclusion of these instances, 39 observations of *reuen* remain.

4.2.3.2. Variables

4.2.3.2.1. Construction Type

As would be expected from their predominantly early attestation, the majority of observations of *reuen* are constructed impersonally with an oblique experiencer argument. Specifically, this applies to 29 observations, all of which were identified by the case marking of a pronominal experiencer, while 7 observations, which either contained an experiencer realised by a full NP, the uninflected relativiser *þe*, or a null constituent, remained ambiguous. This leaves a total of 3 personal constructions, 2 of which contained an experiencer marked by nominative case and 1 of which was analysed on the basis of verb agreement. The diachronic distribution of the 3 personal constructions is not particularly revealing, being spread out across all three periods that contain unambiguous observations. The earliest occurrence is in the *Trinity Homilies*, whose manuscript is dated to M1, while the 2 remaining instances derive from texts belonging to M3 and M4. All of these periods also contain impersonal constructions, whose frequency declines from 27 instances in M1 to 2 instances each in M3 and M4. Incidentally, one of the impersonal constructions takes the form of a double-object construction, in which both the experiencer and the stimulus argument are marked by oblique case, cf. *me rewep Marie þi sone and þe*. ‘I pity, Mary, thy son and thee’ (CMEDVERN,257.723), also cited in (100) above. While the overall figures display a trend towards a higher proportion of personal constructions, the extremely low frequencies do not permit any definite statement beyond the assertion of a virtual disappearance of unambiguous observations of *reuen* after M1.

4.2.3.2.2. Semantic and Syntactic Properties

The majority of the attested instances of *reuen* are constructed with a non-human stimulus argument, which usually invokes the subsense of regret or repentance, but which can also occur in contexts in which pity towards a particular aspect of another human’s condition is expressed,

cf. e.g. *[m]ildheorted beð þe man þe reouð his nehgebures unselðe . [...]* ‘merciful is the man who pities his neighbour’s misfortune’ (CMTRINIT,95.1288). Other typical stimulus arguments include *hise sennes* ‘his sins’ (e.g. CMVICES1,121.1497), but also complex propositional stimuli such as *[...] I haue not lovyd þe alle þe days of my lyue , & þat sor rewyth me ;* ‘I have not loved thee all the days of my life, and that grieves me sorely’ (CMKEMPE,50.1137–1138). The propositional stimuli generally refer to past situations that the experiencer regrets rather than to hypothetical actions requested of the experiencer, and they are thus quite different from the frequent type of propositions observed with *liken* and *plesen*. Human stimuli, on the other hand, which include *mann* ‘man’ (e.g. CMORM,INTR.L57.112) and anthropomorphic participants like *enngleflocc* ‘flock of angels’ (e.g. CMORM,INTR.L57.113) are less frequent and generally representative of the specifically religious nature of the data. Not all of these instances necessarily imply compassion for another human participant, who may also simply be regarded as the source of grief as in | *Himm reoweþþ ec off alle þa* | | *Þatt follþhenn deofless lare ,* | | ‘he also feels pity for / is grieved by all those that follow the devil’s lore’ (CMORM,I,193.1583). Regarding the categorial realisation of the stimulus, the frequent co-occurrence of PPs with human stimuli may be worth noting. These are most prominent in the *Ormulum* (cf. the previous example), but they are also attested twice in later texts, cf. e.g. *[...] ' I haue rewþe on þe peple* , ‘I have felt pity for the people’ (CMWYCSE,248.442) in the *English Wycliffite sermons* from M3.

4.2.3.2.3. Discussion

The limited attestation of *reuen* and the dominance of impersonal constructions make it difficult to draw conclusions about the potential influence of the different types of stimuli on the development of the construction type. All of the 3 personal constructions that occur in the data are constructed with human stimuli, which may indicate an association of personal constructions and animate participants, especially in view of the fact that the majority of observations involve non-human stimuli, which makes their failure to be attested in personal constructions less likely to be the result of pure chance. The situation is complicated by the different subsenses of *reuen*, however, which have different potentials for the occurrence with human and non-human participants. Regret or repentance are mainly applicable to abstract moral features or misdeeds, but not so much to human participants. This means that observations exhibiting this subsense, which is prominent in the religious literature of the primary sources, cannot really be expected to vary in terms of the animacy of their stimulus

argument. Human stimuli, on the other hand, qualify for the subsense of pity, although they do not automatically imply a compassionate reaction of the experiencer, who may simply be grieved by another participant's conduct without the dimension of empathy. The motivations behind the overarching sense of grief expressed by *reuen* may thus be quite different, and this impacts the participants' potential of being construed as agentive. Repentance as the result of active contemplation may result in a more agentive interpretation of the experiencer, which would thus be expected to occur with non-human stimuli given the restriction of this subsense. Grief and pity, on the other hand, may be caused by either human or non-human stimuli, and the present hypothesis predicts that the latter have a higher potential for an agentive interpretation. Whether this potential is present regardless of whether the emotion is primarily located in the personal sphere or whether it involves sympathy for another participant is debatable, since the causation of both grief and pity may be compatible with the intentions of a given stimulus. The 3 personal instances in the data appear to be limited to the latter sense, however, and the animacy of their stimulus arguments is obviously not compatible with the present hypothesis, which predicts that agentive contexts of the stimulus argument favour impersonal constructions with oblique case marking of the experiencer. Instead, the personal instances might be connected to the relatively intimate nature of the situations that the sense of pity implies, which would be reminiscent of the late personal instances of *liken*. This is difficult to substantiate on the basis of 3 observations, however, which are also chronologically disjointed and thus not indicative of a diachronic trend in the data. Overall, the observations of *reuen* thus do not offer particularly robust evidence, although the tendency that emerges rather contradicts than confirms the hypothesis.

4.2.4. Longen

ME *longen* 'to desire' is the final verb of emotion to be considered in this study. Similar to *reuen*, it covers a different semantic area than the three near-synonymous verbs of liking that form the core of this investigation, but in contrast to that verb, the emotion of *longen* does not seem to be interpretable as a cognitive act involving volition. The experiencer's potential to be interpreted as relatively more agentive thus rests entirely on the negative interpretation of the stimulus argument in this respect. Non-human and abstract stimuli should favour the use of personal constructions with *longen*, while human and concrete stimuli are expected to continue to be used longer in impersonal constructions.

4.2.4.1. Database

The overall representation of *longen* in the two corpora is somewhat more favourable than that of *reuen*, although the chronological distribution of its occurrences is not particularly even. A complicating factor in the extraction and identification of forms of *longen* is the existence of the homophone verb *longen* ‘to be suitable, belong’ (MED26032), which proved to be particularly frequent in certain collections of letters in the *PCEEC*, especially in M3 and M4. After their exclusion, the resulting dataset of *longen* (MED26030) comprises 197 observations, 23 of which derive from the *PPCME2* and 174 from the *PCEEC*. This difference already betrays a greater tendency of *longen* to occur in the correspondence rather than the religious literature. The absolute frequency of its observations by period and the standardised frequency of *longen* per 100,000 corpus words are given in table 16 below:

Table 16 Absolute and standardised frequency of <i>longen</i> by period.						
						(n = 197)
	M1	M3	M4	E1	E2	E3
absolute	13	5	17	15	88	59
standardised	5.0	1.2	2.2	4.9	9.7	10.6

As can be seen, *longen* is most frequent in the last two periods E2 and E3 with 88 and 59 observations respectively, which correspond to about 9.7 and 10.6 occurrences per 100,000 words. The highest frequency in the *PPCME2* is found in M1 with 13 instances or about 5.0 occurrences per 100,000 words. The remaining observations are distributed across the central periods M3–E1 and comprise 5, 17 and 15 instances respectively, which correspond to about 1.2, 2.2, and 4.9 occurrences per 100,000 words in both corpora combined. The frequency of *longen* is thus generally much lower than that of *liken* or *plesen*, and the relatively late increase in frequency equally contrasts with the development of the latter two verbs.

Of the 197 observations of *longen*, 24 are non-finite instances and 6 are passive constructions. The non-finite instances predominantly consist of present participles, which sometimes function as participial adjunct clauses, cf. e.g. *[s]o , longing to see thee , I rest Your loving , faithfull , constante frend Charles R.* (ORIGIN3,211.021.207). Quite frequently, however, the present participles also occur as attributive adjectives in collocation with *desire* or, less frequently, with other nouns like *love* or *expectation*, cf. e.g. *[m]y wyfe hath a longinge desier for to see yo=u=* , (CORNWAL,118.079.1062). The experiencer of longing is expressed by the possessor *[m]y wife* in this instance, and the non-finite complement clause *for to see*

yo=u= expresses the object of longing, which could be analysed as equivalent to the stimulus argument. Structurally, however, the head noun *desier* corresponds to the experiencer argument of other non-finite instances, and the infinitival clause is a postmodifier. Since the attributive adjective mainly has an emphasising function with regard to the head noun, which already expresses a kind of longing or other emotion, and since there is no syntactic stimulus argument connected to the form of *longen*, these cases were left out of consideration. The passive constructions lack the experiencer argument in some cases, and instead, place the focus of the utterance on the stimulus argument as the object of longing, cf. e.g. *[w]hen our harvest is done we purpose God willing to come to see you and to wayte on you home to Hatfeild, where you are much longed for [...]* (BARRING,91.042.801). After the exclusion of such instances, 167 observations of *longen* in finite active clauses remain.

4.2.4.2. Variables

4.2.4.2.1. Construction Type

The type of construction of finite active clauses of *longen* exhibits a relatively clear diachronic divide between M1 and the later periods. While the 11 unambiguous constructions in M1 are exclusively impersonal, virtually all of the 122 unambiguous constructions from the later periods are personal, with the exception of 2 impersonal instances in M4. These figures exclude ambiguous observations, which amount to a total of 34 or about 20.4% of all instances. These are distributed relatively evenly across the periods, which is notable because the higher frequency of pronominal *you* as the categorial realisation of the experiencer argument in the later periods resulted in an increase of the proportion of ambiguous constructions for some of the other verbs. The distribution of the two construction types of *longen*, excluding ambiguous observations, is presented below in figure 22:

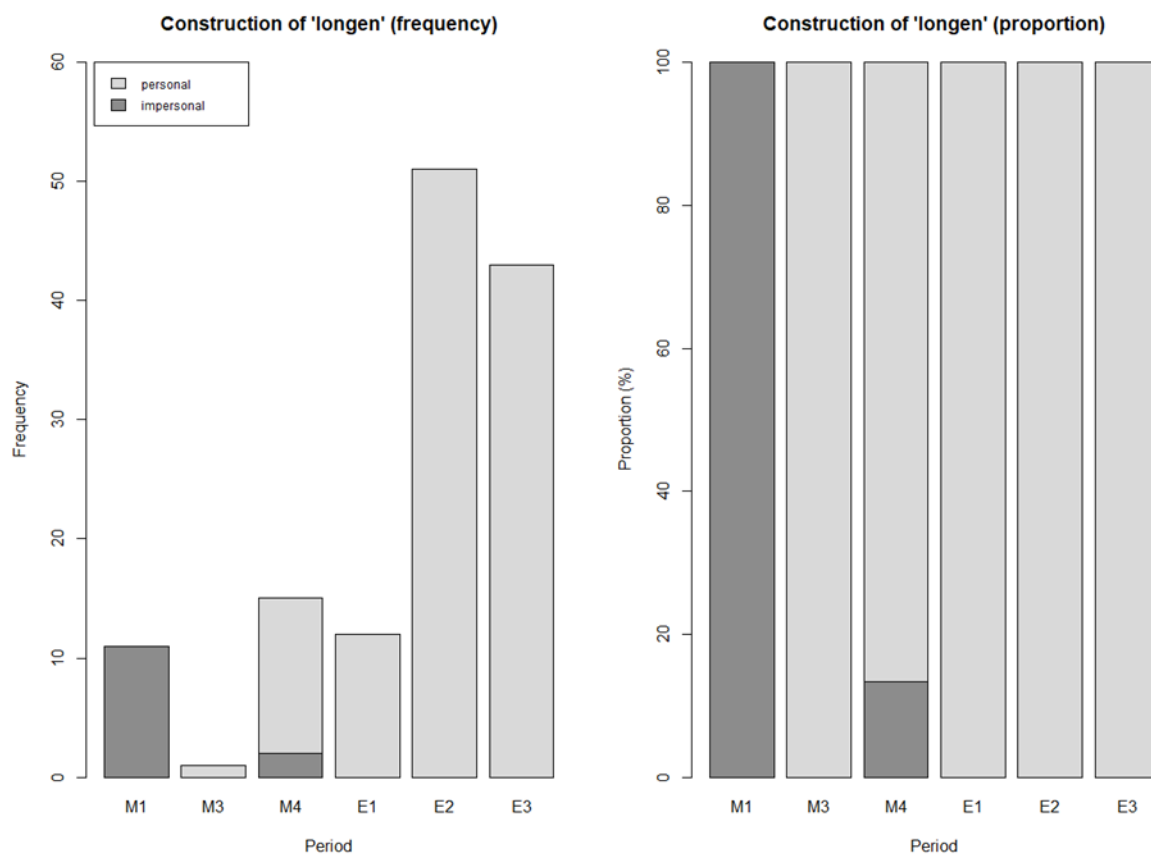


Figure 22 Frequency and proportion of the construction of *longen* ($n = 133$).

Regarding the 2 impersonal observations in M4, it is worth pointing out that both of these derive from the works of Richard Rolle, whose English writings are dated to the end of M2 but included in the *PPCME2* in the form of editions based on later manuscripts. Since all observations were classified in terms of their chronological provenance according to manuscript dates, the risk of potential archaisms is present for all texts whose versions are not based on contemporaneous manuscripts. Since no additional finite active instances of *longen* are recorded in M2, and since the attestation in M3 is equally limited, it is difficult to judge whether the two impersonal instances in Richard Rolle's writings reflect the idiosyncratic use of the author, which may have survived as a characteristic feature in the manuscripts, or whether they are representative of the general use of his time, which may still have been felt to be acceptable at the time of copying, despite the fact that the other observations from M4 are overwhelmingly personal. Either way, these instances highlight the methodological difficulties connected to the investigation of historical sources, which involve repeated copying of their texts. For the purpose of discussion, the data will be taken to reflect a practically complete shift from impersonal to personal constructions after M1/M2.

4.2.4.2.2. Semantic and Syntactic Properties

Most instances of the stimulus argument in the dataset including non-finite instances and passive constructions are non-human, which may be somewhat surprising, since the primary sense of *longen* seems to be favourable towards a more frequent occurrence of human stimuli. Disregarding 9 observations with a missing stimulus argument, 12 or only about 6.4% of all 188 instances comprise human stimuli and 176 or about 93.6% non-human stimuli. The actual difference between the two semantic categories may not be as severe, however. For example, the feeling of longing for another person and the feeling of longing to see another person are semantically quite close. The latter of these expressions is frequent in the correspondence corpus besides similar expressions of longing to hear from, be with or speak with another person, all of which were recorded as semantically complex propositional stimuli, cf. e.g. *[t]hus longing for you more then a thirstie man longeth for a drincke , I bid you hartelie fare well .* (STOCKWE,II,19.062.1200) compared to *[m]ost loving Brother I long to see you ,* (ORIGIN3,96.011.97). The distribution of human stimuli does not display any characteristic diachronic trend, although it is, perhaps, worth noting that none of the instances in M1 involves a human stimulus argument. The same is true of observations in E1, however, which means that the absence of human stimuli is not an exclusive feature of the religious literature in M1, but also characteristic of certain sections of the correspondence corpus.

As with *liken* and *plesen*, the most frequent type of non-human stimuli in terms of abstractness are propositions, but in contrast to the former verbs, their proportion across the different periods is relatively consistent with *longen*, typically ranging between 72.0% (E2) and 86.7% (E1). Two exceptions to this consistency are presented by M3, which contains no propositional stimuli in the 4 total observations, and M4, which contains an above-average proportion of human stimuli (6 or 37.5% of all 16 instances in that period), which, in turn, implies a below average proportion of propositional stimuli (8 or 50% of all instances). The proportions of all semantic types, including human, non-human concrete, abstract, and propositional stimuli, are presented below in figure 23:

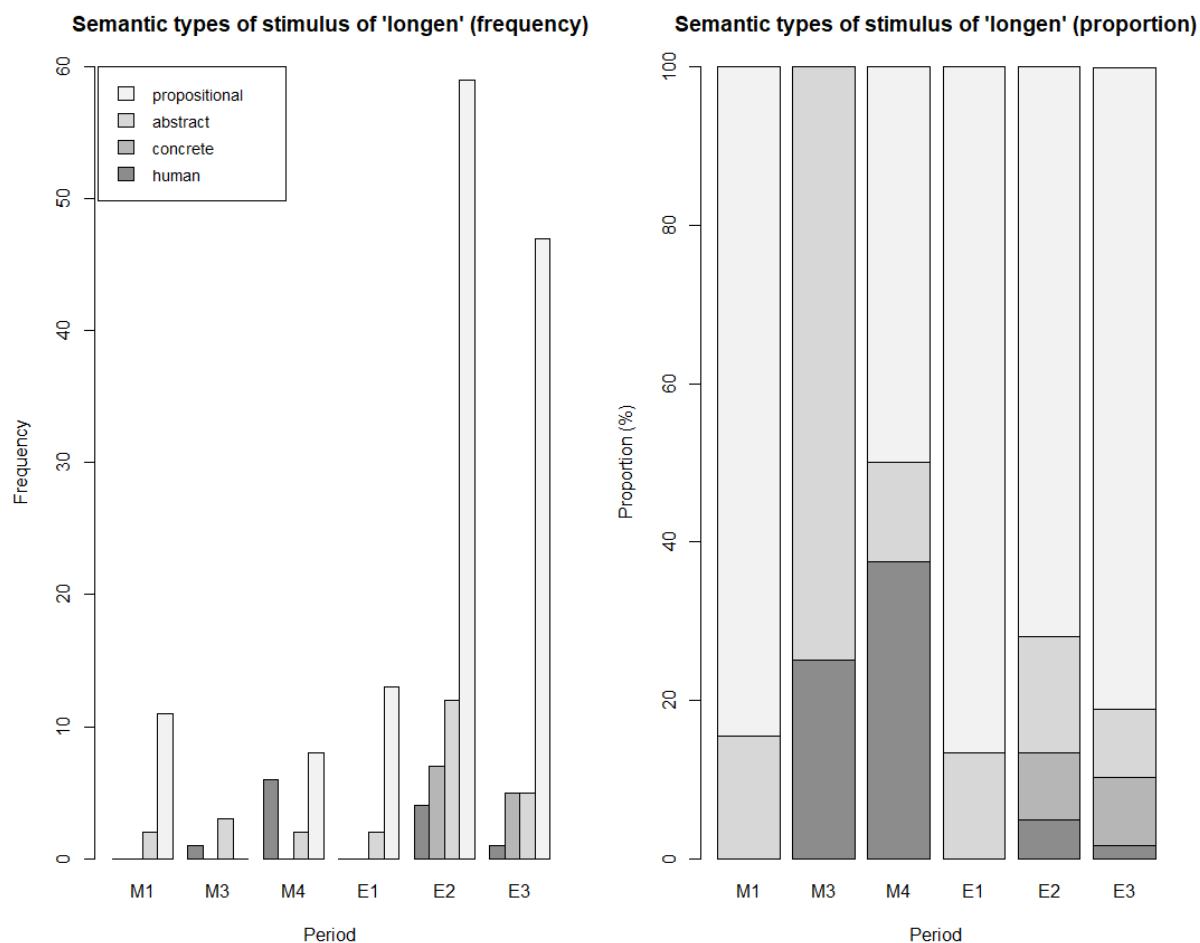


Figure 23 Semantic type of the stimulus argument of *longen* (n = 188).

As can be seen from the figure, concrete stimuli are not very prominent in the data and restricted to the later periods E2 and E3, in which observations of *longen* are also generally more numerous. They include several references to points in time like *that happye daye of setting* (BARRING,73.026.559), but they are not otherwise limited to any conspicuous category. Abstract stimuli are distributed more widely across the periods by comparison and include some of the abstract notions that would be expected in the religious literature, cf. e.g. *blisse* ‘bliss’ (CMANCRIW-2,II.275.492) and *hys lufe* ‘His love’ (CMROLLEP,70.140).

In terms of categorial realisation, the propositional stimuli predominantly take the form of infinitival clauses. Even in 2 instances realised by a finite subordinate clause, the subject of the embedded clause is coreferential with the preceding experiencer argument. Apart from typically containing an explicit rather than an inferred stimulus argument, the constructions with propositional stimuli are comparable to the respective constructions of *plesen* (and *liken*), for which a more volitional interpretation of the experiencer on the basis of their performance of the action in the non-finite subordinate clause has been discussed. This interpretation does not seem to carry over to *longen*, however, although a more agentive subsense ‘to demand,

request' would be conceivable for this verb, in principle, similar to the more agentive sense of *plesen* as conveying a decision. The availability of this implication in a given syntactic structure does not seem to present a sufficient condition, however, which also seems to be connected to the fact that the predicates of the infinitival clauses that complement *longen* typically express less controllable situations, which suit the emotional sense of longing. In addition, it should be noted that a subset of observations was included in the number of propositional stimuli that involves the inference of a verb of movement, cf. e.g. [...] *him wile sone longe þiderward* 'he [sc. the righteous man] will long [to go] there soon' (CMLAMB1,157.500). This type of observation, in which *longen* is constructed with a directional adjunct, is the most frequent one with a propositional stimulus argument in the *PPCME2* and comprises 10 instances compared to only 3 instances with a propositional stimulus argument in the form of an explicit infinitival complement.

Virtually all other instances of *longen* in the two corpora are constructed with PP stimuli, which are most typically headed by *for* or, in some early instances, by *after* or *to*. The only exception to this is a single instance of a stimulus realised by a relativising pronoun in M4, cf. *I it am þat þou langes*. 'it is me that you long for' (CMJULNOR,59.260). This instance is part of an enumerative section in *Julian of Norwich's Revelations of Divine Love*, in which the same sentence is repeated with different verbs in the relative clause. Not all of these verbs appear to conform equally well to the repeated syntactic pattern, and the use of *longen* with an NP stimulus may thus be an instance of literary licence. Regardless of its interpretation, the most dominant types of stimulus arguments are constructions with an infinitival complement and constructions with a PP stimulus. Their chronological distribution can be more or less inferred from figure 23 above, since most of the semantically propositional stimuli are also realised by infinitival clauses, and the other semantic categories are consistently realised by PPs with the single aforementioned exception. The categorial realisation of the stimulus hence does not need to be plotted separately.

4.2.4.3. Discussion

The different semantic and syntactic types of stimuli do not show any prominent associations with either impersonal or personal constructions, which are also difficult to establish due to the clear diachronic divide between the two construction types. Since the unambiguous observations in M1 are universally impersonal, this implies that the probability of impersonal constructions in this period is 100%, regardless of the properties of a given stimulus argument.

In all later periods except M4, the probability of personal constructions is 100% and equally independent of the type of stimulus. Of the two impersonal constructions in M4, 1 is constructed with a PP and 1 with an infinitival complement. Since there are 5 additional personal instances with a PP stimulus in that period and 7 additional instances with an infinitival stimulus, the odds of personal constructions are technically somewhat higher with infinitival complements than with PPs, but these frequencies are, of course, much too low to warrant any further consideration. The consistent realisation of non-propositional stimuli by PPs, which occur from the beginning of the investigated period onwards and thus do not act as a replacement of earlier genitive complements, appears to be a possible obstructing factor in the interpretation of that participant as agentive, since its categorial realisation is not compatible with that of English subjects. The relatively early shift from oblique to nominative experiencers may thus be at least indirectly favoured by the encoding of the stimulus as a non-subject constituent.

With propositional stimuli, a positive argument for the early adoption of nominative experiencers could be made based on the coreference of the experiencer with the PRO subject of the embedded predicate. This line of argument is presented by Allen regarding the NO PROP construction of *plesen*, in which a propositional stimulus is not explicit but inferable from the preceding context. The indirect inference of a more agentive experiencer argument in these constructions appears to be connected to a range of specific clause types, which involve an aspect of indefiniteness regarding the time, place, manner or realisation of a given action, which is at the experiencer's discretion. It is worth noting that these constructions are attested earlier with *liken* than with *plesen*, the former of which is constructed with an oblique experiencer in such instances in M3 (*as him liked*), while comparable instances of *plesen* exhibit an expletive constituent in M4 (*as it pleased him*). Moreover, they do not occur in unambiguous constructions with a nominative experiencer until the later EModE periods, especially in E2–E3. At the same time, the experiencer's performative role in the realisation of a subordinate event can also be implied in cases with an expressed propositional stimulus rather than an inferred stimulus argument. In these cases, the performance of an action by the experiencer of *plesen* is either reported or requested by the speaker, and they do not lead to an early adoption of nominative case but are, instead, typical of the extended preservation of impersonal constructions with an expletive, which are employed in contexts of heightened politeness. With *longen*, the indirect inference of a more agentive role of the experiencer appears to be equally available, in principle, from the construction with infinitival clauses, but the verbs that occur in the infinitival clauses, including *see* and *hear* as the most frequent items, are not typical action verbs and thus appear to be less conducive to such an inference. Instead, they indicate situations

which cannot be immediately realised by the speaker, and they are thus not only characteristic as a way of greeting the addressee of the written correspondence with an expression of longing but also appropriate for the use of *longen* in the literal sense of experiencing a feeling of desire. Even if the inference of a more agentive experiencer based on coreference with the agent of an embedded predicate is available, this inference thus does not automatically take effect for all types of experiencer verbs, and, in view of the relative chronology of *liken* and *plesen*, it does not seem to have had a crucial impact on the extension of nominative case to the experiencer in either case.

Even though the early adoption of nominative case marking for the experiencer of *longen* does not seem to be relatable to an agentive interpretation of the experiencer based on the properties of the stimulus argument, there are other factors that potentially explain the different development of *longen* compared to other verbs of emotion. One notable difference is the verb's relatively low frequency during the middle periods M4–E1 compared to *liken* and *plesen*, which are increasingly used with propositional stimuli and an expletive in the correspondence corpus. Such instances are entirely absent with *longen*, whose frequency does not increase until the later periods E2–E3. This difference in the diachronic distribution of the three verbs appears to be associated with a difference in their dominant communicative function in the correspondence. The verbs of liking are frequently used to utter requests, which require a heightened degree of politeness in formal contexts. This is reflected in their frequently formulaic use in conjunction with abstract terms of reference, which are typically applied as a way of formally addressing second-person participants. The experiencer of *longen*, on the other hand, much more frequently refers to a first-person participant, which is the case in 131 or about 67.2% of all 195 observations in which an experiencer argument is present. Incidentally, this increases the proportion of classifiable constructions in the later periods compared to *plesen*, whose construction is much more frequently ambiguous after the abandonment of the expletive pronoun on account of the higher frequency of morphologically ambiguous *you* and formal NP constituents such as *your Lordship*. The NP constituents that occur with *longen* are also more frequently of a personal nature, cf. e.g. *Howr mother* 'our mother' (CELY,155.117.2623) and *[y]our daughter Lucy* (BARRING,235.177.3050), while abstract terms of address are limited to 2 instances in the entire data. In sum, these facts indicate that the attestations of *longen* are of a more private nature than the frequent use of *liken* and *plesen* in polite contexts. Its increased frequency in E2–E3 may thus also be the result of an increasing representation of private correspondence, which would also agree with the increasing attestation of human stimuli with *liken* in these periods. The characterisation of *longen* as a more private verb is, of course, not

sufficient as an explanation of its early shift to personal constructions, but it does indicate that verbs of emotion can vary significantly in their use depending on the communicative function that they fulfil. While *liken* and, especially, *plesen* exhibit an almost inflationary increase, which culminates in formulaic expressions of politeness, the more literal use of *liken* and *longen* as expressions of emotion is much less frequent overall and, apparently, less suited to the conventions of the correspondence represented in the *PCEEC*.

4.2.5. Wanten

The final verb to be considered in this study is ME *wanten* ‘to lack’. Its foreign origin and secondary integration into the system of oblique-experiencer verbs in Middle English are reminiscent of ME *plesen*, although *wanten* is, of course, of Scandinavian rather than of French origin. In contrast to *plesen*, the verb is represented in the corpora from the earliest ME period onwards and continues in the primary sense of lacking throughout the investigated period, while the typical PDE sense of desire remains largely absent. ME *wanten* is thus semantically distinct from *longen* and also from verbs of emotion in general, since it is classified as a verb of lacking, whose argument structure does not logically presuppose an animate experiencer argument in the narrow sense. The potential variability of the equivalent argument of *wanten* opens up the possibility of investigating the associations of both arguments with the construction type rather than merely those of the stimulus argument in the case of the verbs discussed so far. As with the previous verbs, the following section will first provide a description of the database and the realisation of the variable properties in the observations of *wanten* before offering a discussion of the data.

4.2.5.1. Database

With a total of 447 observations, the attestation of *wanten* is considerably more favourable than that of most of the verbs of emotion that have been discussed so far. Similar to *longen*, the diachronic distribution of these observations is quite uneven, however. Most observations derive from the last two periods E2–E3, which contain about 85.5% of all observations. Their absolute frequency and the standardised frequency of *wanten* per 100,000 corpus words are given for each period in table 17 below. Since M3 and M4 are the only periods in which the *PPCME2* and the *PCEEC* overlap, and since there are generally few observations of *wanten* in

these periods, the two corpora are not distinguished in the table. Instead, the figures in the periods with overlap are based on the total observations from both corpora combined:

	M1	M3	M4	E1	E2	E3
absolute	10	21	13	21	182	200
standardised	3.9	5.0	1.7	6.8	20.0	36.0

The increase of observations in the two final periods E2–E3, similar to *longen*, is notable, and it is even more pronounced than for that verb. There are 182 and 200 observations of *wanten* in these periods respectively, which correspond to about 20.0 and 36.0 instances per 100,000 corpus words, while the highest frequency of *longen* was 10.6 instances per 100,000 words in E3. Whether this increase of *wanten* equally points to a shift towards more private correspondence in the later periods needs to be assessed on the basis of its individual occurrences, however, since it cannot be generally inferred from the meaning of the verb.

The relatively high number of observations of *wanten* is significantly reduced if non-finite instances and observations that are missing at least one of the two arguments are disregarded. The reduced dataset contains 268 observations and is thus about 40.0% smaller than the full set. The reduction is mainly attributable to non-finite instances, which constitute 113 or about 25.3% of all observations, but also the number of observations with a missing experiencer argument is particularly high when compared to other verbs, amounting to a total of 105 instances, 52 of which are part of the non-finite instances and an additional 53 of which constitute finite active clauses. Also the stimulus argument is missing in 20 observations, 6 of which are non-finite instances and 14 of which are additional cases. Among the non-finite instances, present participles constitute the most typical type with 89 instances (c. 78.8%), while infinitival clauses are somewhat less frequent with 24 instances (c. 22.2%). Their combined proportion increases in the late periods E2–E3 from about 0–9.5% in the preceding periods to about 26.9% and 29.5% in E2 and E3 respectively, which means that non-finite instances contribute to the overall increase in the frequency of *wanten* in later periods, but they are not the primary factor, since also the frequency of finite active clauses increases during that time.

The observations with a present participle can be divided further into observations that contain a form of *be* or a raising verb like *seem* or *appear* (62 instances) and observations that do not contain such verbs (27 instances). The verb *be* in the former instances was interpreted

as a copula rather than as a structural auxiliary in progressive constructions, which agrees with the analysis of the present participial of *wanten* as a deverbal adjective form in predicative position. The remaining instances are somewhat more varied and include a small number of complex transitive constructions, in which the interpretation of the participle is equivalent to that of the instances with the copula, but also adverbial clauses, participial relative clauses and adjectives in attributive position. The most prominent subtype among the non-finite instances are thus the copular constructions with a present participle. Two examples of this subtype, one with a human stimulus argument and one with a non-human stimulus argument, are given in (156) and (157) respectively:

(156) I shall not be wanting , as I am bound in duty , to make it my earnest request to God for yo=r= preservation . (HADDOCK,2.001.17)

(157) there shall never bee wanting in mee a ready minde and will to performe any service whattsoever you shall comand iff it lyeth in mee . (OXINDE,I,260.153.2514)

The semantic content of these two instances is similar, but the syntactic structure is different in each case. In the former instance, the human participant *I* is the subject of the complex-intransitive clause containing *wanting*, and the infinitival clause *to make it my earnest request [...]* complements the participial adjective by stating the respect in which the author promises not to be wanting. In the latter instance, a similar promise is expressed by an existential clause with expletive *there* and an extraposed NP *a ready minde and will to performe [...]* as subject of the clause, while the human referent is expressed by a locational PP *in me*. The semantic interpretation of *wanten* in both of these instances is causative, since the associated subject encodes the entity that is absent or lacking and not the entity that experiences or exhibits the lack. The latter of these two participants, which can be referred to as experiencer argument in the wider sense, is analysed as missing in both of the above instances. Semantically, it could be identified with the proposition of the infinitival clause in (156) and the PP in (157). Under a more conservative analysis, these constituents are not interpreted as verbal arguments, however. An argument in favour of this interpretation is the occurrence of explicit PP experiencers in instances like the one in (158) below:

(158) Your praiers and hartie wel wishes have not bin wanting to me upon all occasions , (DUPPA,67.033.648)

In these instances, the entity that experiences lack, or, as is frequently the case, does not, is explicitly expressed by the PP *to me*. It seems reasonable to assume that either a specific human

experiencer, most likely the addressee, is to be understood in (156) as well, even though the phrasing is that of a general statement like the one with existential *there* in (157). The interpretation of the verbal meaning as causative is somewhat less straightforward with human stimuli, since human participants qualify more readily for the role of experiencer than that of the stimulus. The human participant in *I shall not be wanting [...]* in (156) could thus also be interpreted as experiencing lack, which would equally result in a kind of deficiency that could be experienced by another participant. The parallel of non-human stimuli in this construction suggests, however, that the causative interpretation is also applicable to human participants. These instances of participial adjectives of *wanten* thus generally contrast with present participles in adjunct clauses, whose subject typically expresses the entity that experiences lack and therefore implies a receptive interpretation of the verb.

4.2.5.2. Variables

4.2.5.2.1. Construction Type

Of the 268 finite active instances of *wanten* with two arguments, 86 or about 32.1% are ambiguous with regard to their construction type, which was established based on the case marking and syntactic realisation of the experiencer, control of verb agreement, and, in some cases, on the basis of the case-identity constraint, while constituent order in instances with an expletive did not play a role as a criterion. The relatively high proportion of ambiguous instances is not unexpected given that most observations derive from the later periods, in which instances of the second-person pronoun *you* are typically not distinctive. Third-person experiencers are even more frequent in ambiguous observations, however, either as deleted constituents or as full NPs or uninflected pronouns. The remaining 182 observations were analysed as impersonal according to the above criteria in 18 instances (c. 9.9%) and as personal in 164 instances (c. 90.1%). Their distribution across the different periods is given in figure 24 below, which displays the absolute frequencies in the left-hand plot and the proportions in the right-hand plot:

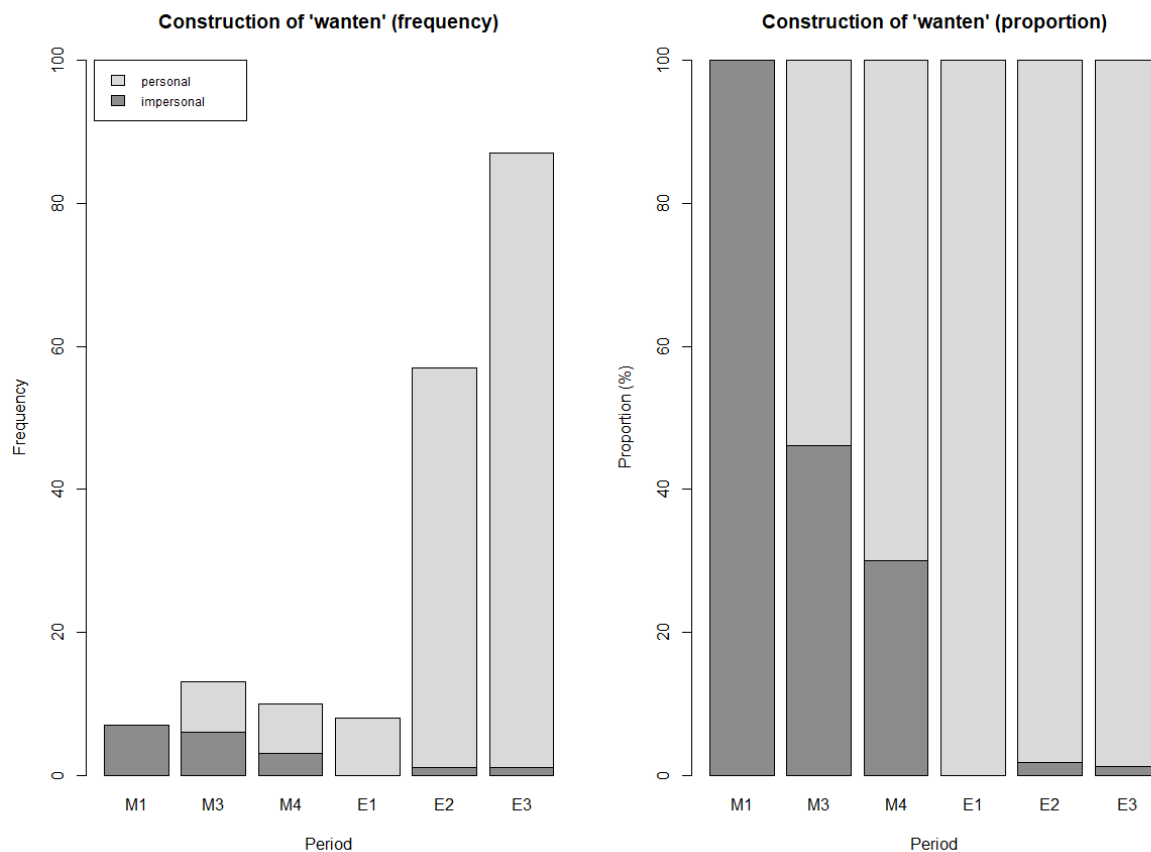


Figure 24 Frequency and proportion of the construction of *wanten* (n = 182).

The two construction types exhibit a steady trend towards personal constructions, whose proportion increases from 0.0% in M1 to about 53.8% and 70.0% in M3 and M4 respectively. From E1 onwards, the construction of *wanten* is almost exclusively personal with the exception of 1 instance in E2 and E3 each, which result in proportions of 98.2% and 98.9% respectively. The significance of this trend is, of course, offset by the low number of observations in M1–E1. After the exclusion of non-finite instances and ambiguous constructions, these contain between 7 and 13 observations each and comprise a total of 38 observations or about 20.9% of the dataset, while E2 and E3 combined account for 144 or about 79.1%. The two impersonal constructions in these later periods are noteworthy, since they appear to demonstrate that the shift towards personal constructions is not fully completed by the late EModE period, despite the absence of impersonal constructions from the data in E1. These instances are cited for inspection in (159) and (160) below:

(159) for I know y=t= **most women in my case wantes** not svffring of mvch paine & trovble
: [...] (FERRAR,283.026.518)

(160) There wants no argument **to so good a minde** . (CONWAY,448.088.2560)

The first example, which derives from a letter written by Susanna Mapletoft in 1634, contains an experiencer argument whose head noun *women* is not marked for case. The interpretation of the example as impersonal thus rests on the agreement of the verb form *wantes* with the third-person singular stimulus argument *syffring of mvch paine & trovble* as opposed to the third-person plural experiencer argument. A closer look at the letter from which the example is taken does not indicate any inconsistency in the use of the verbal ending *-s*, although there are admittedly few contrastive forms, as most verbal forms in the letter refer to the first person. However, there are a couple of forms in a letter written in the 1640s by Anna Mapletoft, a relative of the author, which indicate that the verbal ending *-s* may have been neither restricted to nor consistently used with third-person singular constituents at the time. This concerns the verb form *desiers* in *my father and mother desiers mee to Remember them most kinly to you* (FERRAR,297.030.637), which is used with a plural subject *my father and mother* and thus must be either used impersonally itself or indicate the use of the verbal ending *-s* with a third-person plural constituent. Conversely, the form *giue* in *and she [sc. my aunt] giue you many thanks for your Clues* (FERRAR,297.030.636) indicates that the same ending can be missing with a third-person singular constituent. Since neither of these instances are from the same author as the potential example of an impersonal construction above, they need not call the validity of the example into question, but they demonstrate that the criterion of verb agreement rests on the evidence of verbal endings that are not necessarily standardised. The second example derives from a letter written in 1679 by Henry More. It is exceptional in its use of existential *there* in conjunction with a PP experiencer *to so good a minde*, since these are mutually exclusive elsewhere in the data. The realisation of the experiencer argument by a PP is regarded as equivalent to oblique case marking, and the example thus constitutes an instance of impersonal construction according to the present definition, even if the form of an existential clause differs from earlier instances of impersonal constructions of *wanten* with an oblique experiencer argument. The overall decline of the impersonal construction of *wanten* by the EModE period is nevertheless apparent from the data, while the attestations during the ME period suggest that this decline was less rapid than the one that took place with *longen*.

4.2.5.2.2. Semantic and Syntactic Properties

The stimulus argument of *wanten* is predominantly non-human in all of the investigated periods, and the preponderance of non-human stimuli is even somewhat more pronounced in the full dataset including non-finite instances and instances with a missing experiencer argument.

Disregarding those instances whose stimulus argument is either missing or mixed in terms of animacy, only 77 or about 18.1% of the 426 remaining observations contain a human stimulus argument, and 70 of these occur in the late periods E2–E3. In the reduced dataset of unambiguous finite active instances with two arguments, there are 25 human stimuli, which constitute about 13.7% of the total 182 observations. The reduction is due, to a large extent, to the exclusion of copular clauses with the present participle, which exhibit the highest proportion of human stimuli (26 or c. 41.9% of 62 instances). Regarding the nature of the non-human stimuli, the most notable point is the near-absence of propositions, only a single instance of which is recorded in the entire dataset. An overview of the diachronic distribution of the semantic types of the stimulus argument of *wanten* is given in figure 25 below:

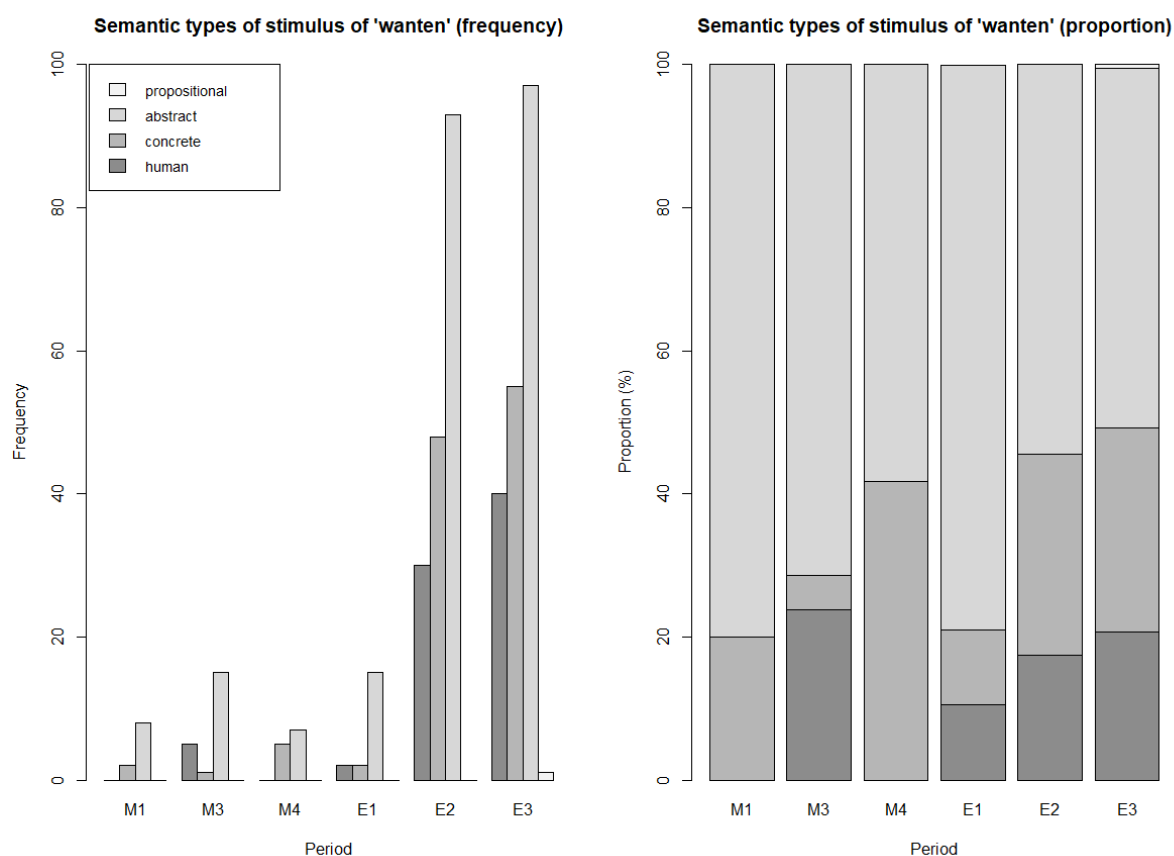


Figure 25 Semantic type of the stimulus argument of *wanten* (n = 426).

The figure shows that abstract stimuli are the most prominent semantic type in all periods, ranging in proportion between 80.0% in M1 and 50.3% in E3. The frequencies between M1–E1 are, of course, relatively low, and the distribution cannot necessarily be interpreted as a reliable diachronic trend. Concrete stimuli, on the other hand, are generally only about half as frequent as abstract stimuli. The latter comprise abstract notions such as *byleue* ‘belief’ (CMWYCSER,333.1910) or *my earnest indeavour* (CORNWAL,39.029.350) and *good advise*

(HUTTON,95.023.293), but also indefinite participants like *any thing* (COSIN,I,2.001.25) or, frequently, *nothing* (e.g. CONWAY,182.035.1092), which were analysed as abstract entities. The relative dominance of abstract stimuli may be somewhat surprising, since concrete stimuli appear to be an equally plausible option with a verb expressing lack. It becomes more easily understandable, however, when the two corpora's dominant genres are considered in more detail. These frequently involve abstract notions in religious contexts as well as abstract qualities that the authors in the correspondence corpus promise to each other. It is also notable that *wanten* is frequently used in negative clauses to express that something will not be lacking or, alternatively, that nothing will be lacking. The third semantic category of human stimuli, on the other hand, frequently refers to specific functions, cf. e.g. *a supply of faithfull able ministers* (WHARTON,11.003.172), which was analysed as human based on the postmodifier, even though the head noun *supply* is, strictly speaking, non-human, or *dragooners* 'dragoons' (ORIGIN3,294.030.308), which represent a specific military unit. It should also be mentioned that 2 of the 5 instances of human stimuli in M3, which is the only ME period in which human stimuli are attested, refer to God, for which the alternative interpretation as a metonymy for faith would be available. In either case, the human stimuli of *wanten* frequently do not exhibit the typical characteristics of agentive participants. A correlate of this observation is the fact that the stimulus argument is more frequently realised by NPs (293 instances or 68.6% of 427 observations that contain a stimulus argument) than by personal (38 or c. 8.9%) or other pronouns (68 or c. 15.9%). As with *longen* above, the distribution of the different categorial realisations of the stimulus argument of *wanten* is not visualised separately.

The experiencer of *wanten*, on the other hand, is more frequently realised by personal pronouns (196 instances or c. 57.3% of 342 observations that contain an experiencer argument) and other pronouns (42 instances or 12.3%) than NPs (40 instances or 11.7%), which suggests that the experiencer more frequently represents given rather than new information compared to the stimulus argument. In terms of animacy, the experiencer exhibits some of the expected variability, although non-human experiencers constitute a clear minority also with *wanten*. Specifically, the experiencer was analysed as non-human in 30 instances and as human in 312 instances. The former instances often involve discussions of literary works, cf. e.g. *That Poem* (PETTY,61.031.836), but also references to more mundane things like *the whole barne* (OXINDE,I,206.128.1900) and animals like *your horce* 'your horse' (BACON,I,56.039.717). More frequently, however, the experiencer argument is missing altogether (105 instances), which means that the proportion of the remaining human experiencers is not quite as pronounced (c. 69.8%) when all observations are considered. The observations without an

explicit experiencer argument can be interpreted as instances of general reference, which may imply both human or non-human experiencers and can thus not be counted in favour of either of the two categories. In terms of abstractness, the non-human experiencers are more frequently concrete (19 instances) than abstract (10 instances), and there is also 1 instance of a semantically propositional experiencer (cf. *this* in FLEMING,224.082.1353, which refers back to the preceding statement). The experiencer of *wanten* is thus not limited in its degree of abstractness as a matter of principle. Overall, the characterisation of *wanten* as a verb of necessity rather than a verb of emotion is reflected in the greater variability of the semantic properties of its experiencer argument, although human referents remain the most prominent type of experiencer with this verb. The association of the different types of arguments with the construction type and the distinctive features of *wanten* compared to verbs of emotion will be discussed in the following section.

4.2.5.2.3. Discussion

The difference in the semantic role of the experiencer argument of *wanten* compared to the more literal experiencer of verbs of emotion is not only reflected in the occurrence of non-human experiencers but also in the different complementation patterns of the verb compared to the items previously discussed. The use of existential *there* and locational adjuncts like the one in (157) above illustrates that the entity of which lack is predicated can be conceptualised and expressed grammatically as a location rather than as an immediate participant, which seems to imply a lowered potential for an agentive interpretation. The implication of the present hypothesis that such constituents receive nominative case later appears to be borne out by the comparison with the respective chronology of *longen*, which is constructed personally almost without exception from M3 onwards, while *wanten* still appears to be in a state of transition during M3 and M4. The robustness of this chronology is, of course, difficult to judge in view of the low number of observations of both of these verbs in the ME period. In addition, the comparison with *liken* does not necessarily support the assumed contrast of verbs of emotion with *wanten*, since impersonal constructions of *liken* are attested with a higher proportion during the transitional periods M3 and M4 than impersonal constructions of *wanten*, even if constructions with an expletive and formulaic instances of *liken* are disregarded. Among the other verbs of emotion, *quemmen* and *reuen*, do not provide any additional indications due to their limited attestation, and *plesen* does not exhibit a general shift to personal constructions.

There is thus no obvious difference between verbs of emotion as a whole and *wanten* as a verb of necessity, although the individual verbs do exhibit rather distinct diachronic profiles.

The limited number of finite active observations of *wanten* compared to the relatively large size of the initial dataset illustrates that a universal treatment of the verb in terms of impersonal and personal constructions does not cover the range of constructions associated with this verb. Similar to the increasing use of passive constructions of *plesen* in E2–E3, the constructions of *wanten* with existential *there* and with a predicative present participle need to be recognised as prominent subtypes, which are part of the verb's general profile. The existential clauses contrast in form with the development of constructions of *liken* and *plesen* involving expletive *it*, but they are comparable to the latter in providing an indirect and hedged way of making reference to an addressee in formal correspondence, since the person affected by potential lack is not explicitly mentioned but implicit in the existential statement. The predicative instances occur in a kind of formula or set phrase with *endeavours*, cf. e.g. [...] *my weake endeavours shall not be wantinge to be a remembrancer [...]* (ARUNDEL,196.030.410), which indicates that *wanten* is susceptible to the stylised formulaic use observable with *liken* and *plesen*, especially in negative statements about abstract qualities, which, according to the author's assurance, will not be lacking. The causative semantics of *wanten* in these participial instances contrast with the receptive meaning observed in non-finite adverbial and relative clauses as well as with the contemporaneous finite clauses, rendering them a semantic archaism, which is even more suited to the formality of the contexts in which these instances occur. The desiderative sense of the verb, on the other hand, which is the central meaning of PDE *want*, does not seem to be required by any of the observations in the data, although instances where either the receptive sense or the desiderative sense would be possible are, of course, present, cf. e.g. *I was with my brother at my Lady Bacon whoe wanteth a dobbel gelding from him & from you a single*. (BACON,II,280.298.5166). This means that the diachronic development of finite active constructions of *wanten* can be discussed in terms of a verb of lacking.

The assessment of the associations of the construction type with different types of arguments is complicated by the uneven realisation of both the dependent and the potential predictor variables. The small number of impersonal constructions and the small number of observations in the ME periods in general make any reliable statements about their association with either human or non-human experiencer and stimulus arguments very difficult. In addition, both non-human experiencer arguments and human stimulus arguments are extremely rare in the reduced dataset and occur almost exclusively in the late EModE periods, during which their construction is no longer distinctive due to the virtual generalisation of personal constructions.

A preliminary regression model of the odds of the construction type with the animacy of the experiencer as the sole conceptual predictor and a diachronic control variable that includes the conflated ME period and the conflated EModE period as variable levels indicates a stronger preference of human experiencers for personal constructions (log odds ratio 1.0562), although this effect does not meet the desired significance level ($p = 0.514$). It appears to be largely based on the much higher number of observations with a human experiencer in the EModE period compared to observations with non-human experiencers, which results in a greater proportion of personal constructions with the former than with the latter (99.3% compared to 85.7%), while each type of experiencer argument occurs only once in impersonal constructions in this period. An equivalent model with the animacy of the stimulus argument as the sole predictor indicates that human stimuli decrease the log odds of personal constructions by a log odds ratio of -0.1196, which is, of course, a very minor effect and could easily be the result of chance given the limitations of the dataset ($p = 0.908$). The instability of the coefficient is also apparent from the fact that a more fine-grained configuration of the diachronic control variable with an additional distinction between M1–M3 as the reference level and M4 as one of the two compared levels leads to a reversal of the effect (log odds ratio 0.2934, $p = 0.7922$). Essentially, the data thus do not indicate any particular association of human or non-human stimuli with either impersonal or personal constructions, which could, of course, be simply due to the uneven distribution of human stimuli. Since the human stimuli that occur in the later EModE periods frequently refer to specific functions, it is also questionable whether these would exhibit a distinctive association with either construction type if the construction of *wanten* exhibited a greater variability in these periods, and since the few human stimuli that occur in the ME periods are at least partly interpretable as metonymic references to faith, it is equally uncertain whether a greater sample from these periods would yield a much higher number of animate stimuli that are more readily interpretable as agentive participants in the predication of lack. Similarly weak results are obtained with abstractness of the stimulus as the wanton sole conceptual predictor. The most robust effect is observed when the association of abstract and propositional abstract stimuli compared to (human and non-human) concrete stimuli is measured in combination with a diachronic control variable that contains the ME period as its reference level and the EModE period as compared level, but, even so, the effect falls just about short of statistical significance (log odds ratio -1.1771, $p = 0.133$). At the same time, this configuration of the diachronic variable implies a low level of precision, which is unable to control for the fact that more abstract than concrete stimuli are attested in the ME period and thus have a higher probability of occurring in impersonal constructions on account of the nature of their diachronic

distribution. While ME *wanten* thus presents an interesting example of the diachronic change from impersonal to personal constructions, which appears to be fully documented over the period under investigation, as well as of the more varied patterns that impersonal verbs may enter, the associations of the semantic properties of its arguments do not lend themselves either to support or refutation of the hypothesis underlying this study. The associations of the different categorial realisations, which virtually lack the category of propositions that is dominant with some of the other verbs, were not tested, since the most frequent category of NPs (c. 69.2% in the reduced set) leaves little room for significant contrasts with the other categories, the second-most frequent of which involves non-personal pronouns (c. 17.0%).

4.3. Summary and Global Discussion

The individual presentation and discussion of the data in the preceding sections has shown that the verbs under investigation exhibit extremely disparate developments during the history of English in a number of respects. Specifically, these include the extent of their attestation during the investigated periods, their evolvement from impersonal to personal verbs, their development of constructions with an expletive, their occurrence in alternative clause types, the distribution of the semantic properties of their arguments, and their propensity towards complementation by different syntactic categories. Table 18 below presents an overview of this development, in which the first period in which personal constructions are attested with a higher proportion than impersonal constructions are indicated for each verb, including an additional distinction of constructions without an expletive for *liken* and *plesen*:

Table 18 Overview of the diachronic development of the construction type.							
	M1	M2	M3	M4	E1	E2	E3
liken						x	
liken_w/o_it					x		
quemen							
plesen							
plesen_w/o_it						x	
reuen							
longen			x				
wanten			(x)				

Cells that are shaded in light-grey in the above table indicate periods in which the construction of a given verb is impersonal with a proportion above 50%, and cells shaded in dark-grey indicate periods in which the construction is personal with a proportion above 50%, the chronologically first of these being additionally highlighted by “x” (or “(x)” in the case of near-equal proportions). White cells indicate periods in which no instances of a given verb in ambiguous finite active clauses with two arguments are attested, which is coextensive with the overall attestation of the verbs except for *reuen*, which exhibits 1 additional instance outside these contexts in E2. Several further details about the attestation of the different verbs and their predominant occurrence in impersonal or personal constructions should be noted before attempting a general summary. First, the table includes no indication of a change towards personal constructions for *quemen* and *reuen*, which is correct, since *quemen* is exclusively and *reuen* predominantly attested in impersonal constructions. The attestation is limited to M1 and M2 for *quemen*, however, after which there are no more records of the verb in the data, and it is extremely sparse for *reuen*, which only contains 2 impersonal observations and 1 personal observation in each of M3 and M4. Conversely, no instances of *plesen* are recorded in the earliest period M1, while the construction of this verb equally remains predominantly impersonal throughout the attested timespan if all observations are considered collectively. Since the number of unambiguous observations of *plesen* is relatively low in E3, the preponderance of impersonal constructions, which is not very pronounced in that period with 52.0% compared to 48.0%, rests on a low number of total observations (25 instances). For constructions of *plesen* without an expletive, the number of unambiguous observations is already low from E1 onwards, and the moderately higher proportion of personal constructions in E2 and E3 (68.6% and 66.7% respectively) equally rests on a low number of total observations (16 and 18 instances respectively). For the verbs *longen* and *wanten*, the opposite is true with regard to the number of unambiguous observations, which are generally few before the late periods E2 and E3. For *longen*, no instances are recorded in M2, which means that positive indications of the verb’s shift to personal constructions are only present in M3, which contains a single unambiguous observation, while positive indications of the conservative impersonal construction are already absent after the conclusion of M1. For *wanten*, there are equally no records in M2, but the subsequent period M3 exhibits a less pronounced preponderance of personal constructions (53.8%), continued by a relatively moderate increase in M4 (70%). In combination, these proportions contrast with the limited evidence of *longen*, especially since the only 2 later impersonal constructions of *longen* in M4 may be dependent on the earlier time of composition of the text from which they derive.

Despite the intricacies of the limitations of the data outlined above, table 18 provides an instructive overview of the verbs' development in terms of construction type. Three of the six verbs that were investigated develop personal constructions with a nominative experiencer argument as their default option, viz. *liken*, *longen* and *wanten*. Of these, *longen* is the first verb to exhibit a full shift towards personal constructions, which is virtually complete after M1 according to the limited evidence in the earlier periods. The second verb in the relative chronology is *wanten*, since its development towards personal constructions is less rapid than that of *longen* and not fully complete before the beginning of the EModE period. The verb *liken*, which was studied most thoroughly, is attested in predominantly personal constructions much later than the other two verbs if all observations are considered collectively. Within this set, the first period in which the proportion of personal constructions exceeds that of impersonal constructions is E2 with 86.6%. If observations with an expletive are excluded, however, the turning point in the construction of *liken* shifts to the earlier period E1, in which 86.2% of the relevant observations are already constructed personally. The turning point shifts to an even earlier period if, in addition to constructions with an expletive, observations that constitute formulaic expressions according to the definition of formula 1 detailed in (94) above are excluded, since this formula involved the presence of an expletive only optionally, and since all instances of the formula were analysed as impersonal constructions. Without these instances, personal constructions of *liken* exceed impersonal constructions already in M4 with 54.8% compared to 45.2%. The discrepancy in development of the three verbs is thus reduced if constructions with an expletive and formulaic instances are distinguished, although the overall relative chronology does not change. The impact of constructions with an expletive is also noticeable with *plesen*, which, besides *quemmen* and *reuen*, presents one of the three verbs that do not undergo a general change towards predominantly personal constructions. While *plesen* never develops personal constructions as the default option with all clause types, the decline of the use of an expletive with this verb in the later periods and the attendant rise in the proportion of structurally ambiguous observations results in a higher proportion of personal constructions in the remaining subset of unambiguous observations, but only if observations with an expletive are excluded, as these are highly associated with impersonal constructions also in the few remaining cases. The construction of *plesen* with a nominative experiencer is typically restricted to inferred propositional stimuli in contexts of indefiniteness (e.g. *as you please*), which continues into Present-Day English. The verb *quemmen*, on the other hand, becomes obsolete after M2 and remains an oblique-experiencer verb throughout its attestation, while the evidence of *reuen* after M1 is inconclusive due to the low number of observations.

In addition to the development of finite active clauses summarised above, non-finite instances and passive clauses are of varying importance for the verbs under discussion. For *liken*, occasional instances in the form of a present participle in predicative or attributive position (e.g. *leazinges likinde* ‘flattering lies’) and the relatively infrequent occurrence of passive clauses, which are virtually restricted to E2 for *liken*, do not change its overall characterisation as a verb that occurs predominantly in finite active constructions. For *quemen*, on the other hand, its frequent occurrence in the form of infinitival clauses functioning as purpose adjuncts is notable. It seems most readily interpretable as an indication of the didactic nature of the homiletic literature in which the verb occurs, since a prominent concern of this genre are instructions of the reader with regard to behaviour that is in compliance with God’s will, and *quemen* appears to have been an adequate expression of this sense of gratification. A notable feature of *plesen* is the fact that passive/adjectival constructions increase significantly in their proportion in the late periods E2 and E3, which is the same time around which the construction of the same verb with an expletive declines. The effect of the passive construction is that it permits the assignment of nominative case to a preverbal constituent while maintaining the typical assignment of oblique case marking to the experiencer argument in the corresponding active construction. The incentive for an increased use of passive constructions seems to be the organisation of information structure, which is, of course, not specific to the use of passive constructions with oblique-experiencer verbs, but which is additionally motivated by the positional restrictions of clausal constituents and PPs, which do not or not typically occur in preverbal position, and which are the only syntactic types of stimulus arguments evidenced with *plesen* at this stage apart from instances in which such an argument is entirely lacking. The participants expressed by these constituents, in both cases, represent the utterance focus, while the experiencer of *plesen* represents the topic, whose tendency to occur earlier than the focus is realised by the adaptation of the experiencer argument as nominative subject of a passive construction. The apparent requirement of nominative case for constituents in preverbal position and the decline of constructions of *plesen* with an expletive, which equally ensure the relative position of a topical experiencer and a focused propositional argument, are not explained by these observations. Nevertheless, it is notable that passive constructions involving human stimuli, which would qualify more readily as utterance topic, do not occur apart from a single instance in which a particular social function is expressed. Finally, the verb *wanten* exhibits two additional constructions of interest, viz. the construction with existential *there* and the construction with a predicative present participle, which partly overlap in different variations of the expression *there shall not be wanting*. The former of these constructions is not

very frequent overall, but it is notable for the associated lack of an explicit experiencer argument, which, itself, is related to the fact that the experiencer argument of *wanten* does not constitute an experiencer in the narrow sense but refers to a less restricted range of participants which can more easily be left implicit or, alternatively, be expressed in the form of a locational adjunct. The construction with a predicative present participle is particularly frequent in the final periods E2 and E3 and notable for the causative interpretation of its semantics in conjunction with human referents, since this contrasts with the receptive interpretation that is otherwise typical of finite active constructions of the time. Collectively, the aforementioned cases of non-finite instances and passive clauses illustrate that the development of the verbs under discussion cannot be characterised exhaustively without reference to syntactic patterns that are beyond the definition of impersonal and personal constructions.

Within the spectrum of semantic categories of the stimulus argument that were distinguished, animacy, or, more specifically, humanness, presents one of the two endpoints whose association with the construction types was compared to the reference group of non-human non-propositional stimuli. As a predictor of the choice between impersonal and personal constructions, animacy proves to be of fairly limited value, at least if it is regarded as operationalisation of a given participant's potential for identification with the prototypical role of agent. The reasons for this are multiple and include, first of all, the particularly limited amount of available data for human stimuli. With all of the verbs under investigation except *quemen* and, to a lesser extent, *reuen*, non-human stimulus arguments are much more frequent than human stimulus arguments. In addition to their quantitative limitation, the distribution of human stimuli is often diachronically skewed, which leaves even fewer observations in some of the periods that were distinguished. With *liken*, for example, human stimuli are notably infrequent in the middle periods after M1 and M2, and they only gradually become more prominent towards the end of the investigated period. With *plesen*, human stimuli are attested fairly well in the ME periods, but they are virtually absent in the EModE periods, particularly in finite active clauses. Also *wanten* exhibits a relative increase in human stimuli in the later periods, but these often refer to humans in specific social functions. Regarding the reason for the uneven distribution of the different semantic types of stimuli, it is difficult to establish to what extent this is a result of the specific compilation and combination of the corpora on which the present investigation is based or interpretable in terms of diachronically changing lexical properties. For *liken*, the latter interpretation in terms of a split into a desemanticised modal marker of politeness and a lexical verb expressing the more literal sense of emotion can be considered. At the same time, it needs to be assumed that the religious nature of the data in the

early periods and the correspondence of the later periods each have an impact on the type of observations that were retrieved. For example, the typical association of *quemen* with manifestations of God as the most common type of experiencer indicates that the full range of the verb is not represented in the corpus, and this is also suggested by the fact that the majority of observations derive from a single text. The impact of the dominant type of genre on the distribution of different types of stimuli is most apparent in the contrast between the *PPCME2* and the *PCEEC* regarding observations of *plesen* in M4, which, in the case of human stimuli, derive almost exclusively from the former corpus, while human stimuli are virtually absent in the later EModE periods that are covered exclusively by the *PCEEC*. At the same time, the number of human stimuli increases slightly in the final period E3 if all clause types are considered, and this, in turn, can be interpreted as a corpus-internal change towards more private correspondence, which also appears to be suggested by the overall increase in observations of *longen* in the later periods given the more private nature of the verb. It seems, then, that the re-emergence of human stimuli with *liken* towards the end of the investigated period coincides with a change in the dominant subject matter of the correspondence, a change that is equally observable with other verbs and which suggests that the existence of *liken* as a literal verb of emotion can be presupposed for the earlier periods in which it is eclipsed by the much more frequent and often formulaic polite use, which is equally characteristic of the even more popular use of *plesen*.

Apart from the quantitative and distributional limitations of the attestation of human stimuli, the adequacy of animacy as an operationalisation of a participant's greater potential for an agentive interpretation is challenged by the fact that the semantic hypothesis based on this criterion is generally not borne out by the data. In terms of pure potential, human or animate participants are certainly more readily interpretable as agents, since they intrinsically possess the proto-agent property of sentience/perception and are also typically capable of the other properties of volition and causation that were discussed. The presence of this potential is not necessarily sufficient to result in a measurable impact on the choice between the two types of constructions across all verbs and their various subsenses, however. For example, the implication of agentivity is not necessarily present in the case of human participants that occur as stimulus argument of *reuen* when it denotes the sense of pity or empathy, since the relevant participant need not have volitionally caused or even be aware of the experiencer's sentiment. The basic potential for an agentive interpretation is, of course, present, but the impact of the presumably less frequent situations where such an interpretation actually applies on other instances of human stimuli where it does not apply may be too insignificant to be detectable,

even if a much larger set of observations were available. As it stands, the 3 observations of personal constructions of *reuen* do not exhibit the hypothesised impact of a higher potential for an agentive interpretation of the stimulus argument, since this argument refers to human participants in all three cases, which should result in the selection of impersonal constructions. Also with *liken*, *longen* and *wanten*, the implication of volition or causation is only potentially present and not necessarily applicable to all instance of human stimuli. While the intrinsic likelihood of volitional causation appears to be somewhat higher in the case of *liken*, the instances of human stimuli of *wanten* frequently refer to more general social functions rather than specific individuals, and they are thus less amenable to the kind of agentive interpretation envisaged in the present context. This could, of course, be addressed by an adjustment of the predictor variable in terms of an identification of these instances with non-human stimuli, but the remaining quantitative evidence of human stimuli would certainly not be substantial enough to be informative. For *longen*, the virtual instantaneity of the shift to personal constructions after M1 makes any association of the conspicuously infrequent human stimuli with the construction type uninterpretable. An effect of the potentially agentive interpretation of human participants may be conceivable for this verb from a semantic point of view, but this is countered by the typical categorial realisation of the stimulus argument by either infinitival complements or PPs, neither of which are particularly suited as representations of agentive participants. The realisation by PPs, in particular, appears to convey a notion of spatial distance, which is also reflected in the reoccurring construction of *longen* with a directional adjunct, and it suggests that no immediate transfer between the two participants, which would otherwise favour the interpretation of the stimulus argument as volitionally involved, is possible.

The most readily interpretable association of human stimuli with either one of the two construction types is found with *liken*, and it is the reverse of the hypothesised effect, since human stimuli increase rather than decrease the odds of personal constructions in the fitted regression model when compared to the reference level. The finding also contrasts with early instances of human stimuli with *liken* in M1 and the evidence of human stimuli with *quemem* in M1 and M2. For the latter, an interpretation as volitional cause is generally available and even suggested by the verb's occurrence in non-finite purpose clauses and the conjunction with instrumental adjuncts and pre-modal verbs such as *willen* 'to want' and *mouen* 'may', which indicate volition and ability respectively (cf. also Allen 1995: 147, Trips/Stein 2019: 256). The relevant instances of *liken* are generally comparable to those of *quemem* in this respect, but the interpretation of these early instances does not carry over to later periods, in which the volitional interpretation of human stimuli is much less prominent. The reason for this appears to be that

the contexts in which the human stimuli occur are not identical in both cases. The early instances of *liken* can be regarded as imitations of the use of *quemēn* in religious or homiletic contexts, for which *quemēn* appears to have been the better-suited verb if its particularly frequent attestation in the *Ormulum* is not misleading in this respect. The similarity of the use of these two verbs is also suggested by the fact that manifestations of God are the most frequent type of experiencer argument in both cases, which may not seem particularly revealing in the context of the religious literature that dominates the earliest sections of the corpus, but which indicates, nevertheless, that these instances do not describe the default situation of emotion involving two human participants. Moreover, it is distinct from the less pronounced frequency of *God* as experiencer argument with non-human stimuli of *liken* in the same period. The later human instances of the stimulus argument, on the other hand, are much more readily interpretable as involved in the expression of an emotional stance of the experiencer towards that participant. The equivalent function of *liken* in the middle periods is not observable, which, again, may be interpreted as a gap in the data resulting from the absence of relevant contexts from the prominent religious genres of the corpus, in which case the existence of a change from oblique to nominative experiencers could still be assumed. Alternatively, the lack of attestation could be interpreted as a diachronic change in the lexical properties of *liken*, in which case the question arises which verb fulfilled the basic function of expressing favourable emotion between two human participants at the time.

If the sense of gratification that is prominent with *quemēn* is viewed as distinct from the expression of emotion associated with *liken*, the fact that the more volitional interpretation of human stimuli with *quemēn* does not transfer to the late instances of human stimuli with *liken* is understandable. At the same time, the involvement of the stimulus argument of *quemēn* in an act of gratification renders this participant not only more agentive in relative terms but also makes it more suitable as the conceptual starting point of the situation, which makes the adoption of nominative case marking for the experiencer argument less likely in view of its less frequent occurrence in preverbal position. The failure of *quemēn* to develop into a personal verb is, of course, not contrastively evidenced by the data due to the verb's early obsolescence, but the basic difference in lexical semantics seems to be equally true for *plesen*, which, initially, continues the use of human stimuli that was previously typical of *quemēn* and only later exhibits the prominent increase in propositional stimuli typical of *liken*, eventually surpassing the latter verb in popularity. The contrasting verbs *liken* on the one hand and *quemēn* and *plesen* on the other are therefore not simply two sides of the same coin but constitute lexicalisations of semantically distinct concepts, where pleasing another participant does not necessarily entail

that they like you, and liking another participant does not necessarily entail that they please you. While this may seem like an obvious observation, the early occurrence of all three near-synonymous verbs with human stimuli that are susceptible to an agentive interpretation in the relevant context and the imprecise definition of the experiencer as a semantic cover term that does not express semantic nuances appear to have led to the undue expectation of agentive features for human stimuli also in the case of a more literal sense of emotion between two human participants. Instead, it seems appropriate to recognise a semantic difference between the contrasting verbs that is also reflected in Allen's designation of *quemen* and *plesen* as Dative Object verbs as opposed to the type-ii verb *liken* (Allen's type I). Since the grammatical status of oblique constituents is not of primary concern for the present discussion, the identification of the former verbs as favouring a stimulus topic and of *liken* as favouring an experiencer topic can be given preference, and this inference of the arguments' typical role in information structure based on the difference in lexical semantics can serve as a generalised explanation of the verbs' differing diachronic behaviour.

The only context in which *plesen* develops personal constructions is represented by inferred propositional stimuli that involve an aspect of indefiniteness. This exceptional behaviour is in formal agreement with the semantic hypothesis, since propositions are regarded as abstract and therefore less conducive to an agentive interpretation of the stimulus argument, which, in turn, should make this option more readily available for the experiencer. However, the development does not seem to be attributable to such an interpretation of the experiencer argument, despite its identification with the subject of the implicitly embedded clause, since the option of such an inference does not affect much earlier instances of comparable constructions of *liken*, which continue to exhibit oblique experiencers, and since the option is less plausible for propositional stimuli of *longen*, which typically contain predicates whose realisation is beyond the control of the coreferential experiencer. Instead, the restriction of nominative experiencers of *plesen* appears to be structurally motivated, since the use of passive/adjectival constructions constitutes the preferred strategy of realising preverbal nominative experiencers in conjunction with explicit propositional stimuli or PPs, which, incidentally, confirms the default interpretation of the stimulus as topic in active clauses. In terms of intra-lexical variation, there is thus no certain indication that propositional stimuli favour the adoption of personal constructions. The only other potential indication of a confirmation of the semantic hypothesis is the more rapid adoption of nominative case for the experiencer argument of *longen* compared to *wanten*, since the former constitutes a verb of emotion, whose experiencer ranks higher in the assumed cline of semantic roles from agent to patient than the equivalent

argument of *wanten*. While the identification of different degrees of potential for an agentive interpretation has proven to be difficult for the distinction of different subtypes of the experiencer argument in terms of cogniser, emoter and perceiver, the expectation of an even lower potential of the equivalent argument of *wanten* appears to be borne out by the relative chronology. However, the observation rests on very limited data in the case of *longen*, and since the relative chronology of the development of *liken*, which constitutes the only other investigated verb of emotion with an observable change to personal constructions, does not support the finding, the possible interpretation of the difference between *longen* and *wanten* in favour of the hypothesis should not be overemphasised.

The strongest association in the data is that between propositional stimuli and impersonal constructions of *liken*, which remains strong regardless of the exact configuration of the reference level of non-propositional stimuli as either including or excluding human or concrete stimuli. With the exception of inferred stimuli in contexts of indefiniteness, a similar association can be established for propositional stimuli of *plesen* on the basis of their frequency in impersonal constructions, even though it is not demonstrable by regression modelling due to the uneven distribution of the different types of stimuli and the relative stability of the construction type in finite active clauses. The other verbs do not provide additional evidence in this respect, since the near-synonymous verb *quemmen* is not constructed with propositional stimuli in the early periods and not attested after M2, the attestation of *reuen* is too sparse after M1 to be conclusive, the shift of the verb *longen* to personal constructions is too rapid to offer differential contexts, and the verb *wanten* is virtually never constructed with propositional stimuli. The observed effect of propositional stimuli on the odds of the construction of *liken* is, of course, strongest if the presence of an expletive is not controlled for and much less pronounced otherwise. It seems plausible to assume that constructions with an expletive are at least potentially distinct from constructions without such a constituent in terms of their mental representation with the language user, even though certain formulaic constructions of *liken* can be conceptualised as a single schema in which the expletive is optional. The theoretical subsumption of constructions with and without an expletive under a single constructional opposition between personal and impersonal constructions based on the case marking of the experiencer argument, on the other hand, can be justified with reference to the general change in the predominant employment of morphological and syntactic means in the history of English and the fact that the association of propositional stimuli with impersonal constructions is not restricted to instances with an expletive or formulaic expressions but involves early instances of mostly inferred propositional stimuli in indefinite contexts as well.

A major point of discussion has been whether the categorial realisation of the stimulus argument should be regarded as a syntactic variable or a semantic variable. The fact that inferred propositional stimuli in indefinite contexts do not exhibit identical associations across all relevant verbs and periods seems to indicate that structural considerations, which are affected by the syntactic means available for a given verb at a given time, do play a role for the construction type, especially since the late instances of such stimuli with *plesen* exhibit the opposite association compared to propositional stimuli with *liken*. At the same time, the fact that the other categorial realisations do not exhibit significant associations with the construction type except in those cases in which they are either coextensive with or defined on the basis of the semantic properties of their referents suggests that semantic features, which are also logically prior to their grammatical expression, are the more important factor. In the case of *liken* and *plesen*, the overriding consideration appears to be pragmatic, however, in that impersonal constructions with propositional stimuli frequently perform the function of added formality or politeness. This is indicated independently by their frequent conjunction with abstract terms of address, which are equally used as a means of adding politeness, as well as the fact that the use of both *liken* and *plesen* extends well beyond instances in which a literal sense of pleasure can be supposed to be present. In these contexts, the lexical meaning of the two verbs appears to be bleached to such an extent that it serves as a purely modal modifier of the propositional content of the stimulus argument. This interpretation is also compatible with the inflationary increase of such constructions in the middle periods M4 and E1, an increase that is even more pronounced and sustained longer with *plesen* and results in a highly conventionalised formulaic device that is represented most fully by its appositive use during the latest stage. The grammatical separation of the experiencer from the verbal predicate by the use of oblique case marking in impersonal constructions adequately supports the aforementioned means, since it constitutes an indirect mode of expressing the experiencer's involvement in the propositional statement, which, in many cases, is hypothetical and merely desired of the experiencer by the author of the correspondence in which these constructions are most typically observed. In many cases, the submitted request itself is conventionalised and directed at the organisation of the discourse by way of politely requesting attention for the subsequent information. In other cases, propositional stimuli involve factual statements, whose modification by the use of *liken* or *plesen* highlights the experiencer's authority in the event and thus equally constitutes a feature of politeness. The semantic complexity of propositions is thus a prerequisite to their occurrence in pragmatically motivated contexts of politeness, since they are suited to the expression of requests and susceptible to additional modification by verbs

of liking as indicators of authority, although this does not seem to be related to a greater or lesser potential of these constituents for an agentive interpretation based on their higher degree of abstractness.

At the other end of the spectrum, human stimuli significantly increase the odds of personal constructions of *liken* when compared to the reference group of non-human non-propositional stimuli. This effect is independent of the strong association of propositional stimuli with impersonal constructions, although it is much more moderate and only attested with marginal significance in the revised version of the regression model that includes the presence of an expletive as one of its predictors. The lower significance is understandable in view of the relatively low frequency of human stimuli, which constitute the semantic group most strongly associated with personal constructions, also in comparison to non-human concrete stimuli, although the latter difference cannot be demonstrated with significance. In contrast to propositional stimuli in impersonal constructions, human stimuli in personal constructions appear to be typical of more private contexts, which lack the aforementioned features of formality and politeness and more frequently refer to a literal sense of emotion, which is simultaneously experienced and expressed by a first-person participant. The private nature of the contexts in which human stimuli become more frequent with *liken* appears to be confirmed by a simultaneous increase in the overall frequency of *longen* during E2 and E3, since *longen* generally expresses a more private or personal sentiment, although a certain tendency of the correspondence towards conventionalised use is observable also with this verb. The similar increase of *wanten* during these periods must be regarded as unrelated or, at least, its connection with more private contexts is not immediately apparent. While the advanced development of *liken* at this stage implies a greater chance of personal constructions in general, the correspondence of authors like Dorothy Osborne/Temple, who wrote towards the end of the EModE period, can nevertheless be taken as representative of the characteristic occurrence of human stimuli in personal constructions, which is attested more sporadically in the periods before, and which contrasts with the highly formalised impersonal constructions involving propositional stimuli that occur in the correspondence of authors like Sir Thomas More. The fact that the pragmatic distinction between impersonal constructions in contexts of politeness and personal constructions in more private contexts is restricted to verbs of liking is, of course, related to their semantic aptitude as markers of politeness. The same quality is not present with *reuen*, *longen* and *wanten*, even though collocational expressions such as *longinge desier* and predicative statements such as *my weake endeavours shall not be wantinge* illustrate that these verbs can equally be adapted to the requirements of conventionalised correspondence.

Regarding the association of human and other types of stimuli, it is unfortunate that the amount of data available for these verbs is not more abundant during the relevant periods of change towards personal constructions.

While the pragmatic difference between impersonal and personal constructions may be restricted to the semantically suitable verbs of liking, their evidence shows that oblique and nominative experiencers were not simply in free variation during the EModE period. For the remaining verbs, a comparable pragmatic difference or a difference in terms of the semantic hypothesis underlying this study is not demonstrable but, strictly speaking, also not refutable based on the limited evidence and the limited diversity of the observations in the most crucial periods of change. It seems doubtful, however, whether an augmented database would be likely to impact the findings in favour of the hypothesis. If the greater potential of animate participants to be identified with the prototypical role of agent only affects a small subset of the relevant instances, the detection of any significant impact on the odds of impersonal and personal constructions would require a much larger amount of data. It is also notable, of course, that the few indications that exist besides the clear counterevidence of *liken* generally fail to bear out the tested hypothesis. A more fundamental reason for this could be that the morphological category that is taken as indicative of a more agentive interpretation of the experiencer argument is successively weakened in this respect. The increasing semantic ambiguity of nominative subjects that results from the extension of nominative marking to less prototypical instances, which include the experiencer arguments under discussion, reduces the prototypical relationship of this category with the semantic role of agent that is assumed to function as a model in the directionality of this process. However, while the association of nominative case marking and the semantic role of agent was never absolute in the history of English, since experiencers of emotion like love and hate and perceivers of visual and auditive stimuli are coded as non-nominative subjects already in Old English, the prototypical association, which results from the consistent assignment of nominative case marking to the relevant argument of action verbs in active clauses, can be postulated even for nominative subjects in Present-Day English. Of course, one-argument verbs like *fall* or *die* do not involve agentive nominative subjects, and the subject of passive clauses betrays the typical reversal of the distribution of semantic roles compared to active clauses that accompanies this construction. Nevertheless, nominative case marking has a differential function in active clauses with two arguments in marking the conceptual starting point of a given utterance, which, in the case of transitive action verbs, is typically the agent. This is also the motivation behind attempts to capture the assignment of case marking in terms of implicational hierarchies such as those employed in

different versions of Case Grammar. The alternative view of nominative subjects as a purely syntactic configuration disregards this functional aspect, which is otherwise pervasively attached to the formal units of language. The assignment of nominative case to constituents that exhibit certain behavioural subject properties like control of coreferential deletion would be vacuous. Instead, the identity of prototypical agents and topical participants that exhibit such syntactic properties is a result of the frequent representation of humans and their actions in human discourse. The emphasis that was placed on semantic roles in the present study in favour of the more controversial definition of impersonal constructions as subjectless thus seems justified, even if the developed hypothesis was not borne out and presumably would not be borne out by a greater amount of available data either.

5. Conclusion

The semantic hypothesis about the extension of nominative case marking to the experiencer argument of oblique-experiencer verbs based on their relative potential for an agentive interpretation has not been confirmed by the present investigation. The expected difference between verbs of emotion and *wanten* as a verb of lacking is not reflected consistently in the relative chronology of the verbs that were investigated, and the observed associations of human stimuli, which were expected to decrease the experiencer's relative potential for an agentive interpretation and thus increase the odds of impersonal constructions, and propositional stimuli, which were expected to increase the experiencer's relative potential for an agentive interpretation and thus decrease the odds of impersonal constructions, are the opposite of those predicted by the hypothesis. Instead, impersonal constructions can be demonstrated to possess a pragmatic function as markers of politeness in the most readily interpretable case of *liken*, in which propositional stimuli are frequently, but not always, used in conjunction with an expletive as part of more or less formulaic expressions that help to organise the discourse, communicate polite requests, or reference the addressee's authority in past events. In these contexts, which are equally found and even more prominent with the increasing use of *plesen*, the lexical meaning of *liken* as a verb of emotion is bleached to the point that it can be regarded as a verbal expression of modality, while the more literal expression of emotion is typical in complementary contexts involving human and other concrete stimuli. The former of these are notably restricted to private correspondence, which appears to be less frequent in the data before the final periods E2 and E3. The restriction of the pragmatic difference to constructions of verbs of liking is a result of their semantic aptitude for use in polite contexts, while no equivalent

results emerge for the remaining verbs, which are generally attested much less frequently in the corpora, despite the fact that most of these verbs also exhibit certain tendencies towards conventionalised use.

Despite existing references in the secondary literature to perceived differences in the agentivity of the arguments of experiencer verbs, whose most important aspect was defined as volitional control over the situation in which they are involved, the systematic application of the derived hypothesis about the diachronic development of these verbs and its operationalisation in terms of independently observable criteria proved difficult, since the manual interpretation of a given observation's context was ruled out as being potentially subjective and unfeasible in the context of a quantitative corpus study, and since the criterion of categorial realisation, which constitutes one of the grammatical properties considered as an indicator of agentivity in Timberlake's original contribution about the directionality of actualisation according to markedness, was difficult to motivate in more explicit theoretical terms. The semantic criterion of animacy and the related criterion of abstractness, while relatively easy to establish without too much subjectivity, only present necessary but not sufficient criteria for the interpretation of a participant as more agentive and presumably lead to much weaker associations than the amount of available data and the diachronically skewed distribution of the variable levels allow to demonstrate with significance. At the same time, the limited indications that do exist beyond the case of *liken* and *plesen* rather contradict than confirm the hypothesis. The main reason that opinions about the impact of an agentive interpretation on the distribution of case marking have been maintained in the context of experiencer verbs appears to be that such a relation is, indeed, suggested by the observable features of constructions of *quemen*, which include not only much more frequent instances of animate stimuli compared to the other verbs, but also other indicators like the embedding in purpose clauses, the conjunction with instrumental adjuncts and the complementation of verbs like *willen* and *mouen*. While this type of usage is also found with a limited number of early instances of *liken* in M1, the specific implication of active gratification or adherence to God's will that is typically expressed in the homiletic literature in which *quemen* is most commonly found does not carry over to the chronologically later instances of human stimuli of *liken* as a verb of emotion. This difference between *liken* and *quemen* in terms of lexical semantics also impacts the standard alignment of their arguments with information structure, in which the experiencer of *liken* and the stimulus of *quemen* respectively constitute the default topic. The same difference is applicable to the distinction of *liken* and *plesen* and can serve as an explanation of the different diachronic behaviour of the two verbs with regard to the

development of personal constructions. The difference is neutralised in contexts of politeness expressed by impersonal constructions involving propositional stimuli, in which the two verbs are largely indistinguishable, although *liken* is ultimately superseded by *plesen* in this function.

Regarding the more general conceptualisation of diachronic language change, the results of the present study reinforce the view that accounts of the extension of nominative case marking to oblique experiencer arguments in purely structural terms are insufficient, since they do not reflect the complexity of the complementation patterns involved, do not account for the gradual development of individual verbs across an extended period of time, and fail to capture the creative use of both traditional and innovative syntactic means to express different communicative functions. The assumption of a grammatically required and thus instantaneous reanalysis of structurally ambiguous constituents in particular is untenable in view of the gradual and lexically diverse realisation of the change. The conditions of ambiguity may be met more frequently than some researchers have assumed due to the positional restrictions of experiencer arguments in conjunction with propositional stimuli and the use of morphologically ambiguous nouns as abstract terms of address even between family members, but there is no apparent chronological coincidence of their increased occurrence and the change towards personal constructions. The amount of ambiguity required to trigger an innovative analysis or to result in an innovative analysis that gains ground and becomes the norm in a given speech community is, of course, impossible to establish with certainty, and early instances of morphologically ambiguous NP experiencers may well have been sufficient for occasional reanalyses in language processing if not necessarily in language acquisition. Regarding the process of extension in terms of gradual disfavouring or the actualisation of an innovative analysis, the main point is whether the optionality of two rivalling constructions is conceptualised as part of core grammar or relegated to a set of grammar-external usage rules. Under a usage-based view of language, the former view is preferable, since there is no theoretical need to assume mental representations by discrete syntactic units or parameters that would make impersonal and personal constructions mutually exclusive, and since the use of these constructions with different pragmatic functions for at least some of the verbs that were investigated illustrates that both the traditional and the innovative construction continue to be employed by language users in a complex system of diachronically shifting communicative needs. The definition of contexts in which the innovative construction is expected to occur earlier or more consistently by reference to general markedness principles is of limited use if these do not entail specific expectations about the directionality of a given language change. The formulation of unmarked contexts based on the prototypical association of nominative case

marking with the semantic role of agent was not confirmed as a mechanism that uniformly governs the development of all of the verbs in question. Instead, it seems preferable to assume more local analogies between individual verbs, since *plezen* arguably follows and eventually supersedes *liken* in the function as a marker of politeness, and since the relatively late shift of *liken* as a verb of emotion towards personal constructions in comparison to *longen* and *wanten* may be indirectly influenced by its parallel use in impersonal constructions in this function. It also needs to be acknowledged that the restriction of the preverbal position to nominative constituents is a more general phenomenon in the history of English and not limited to the experiencer argument of oblique-experience verbs. Regarding the complex factors underlying this development, which have an obvious impact on the permissibility of oblique preverbal experiencers, and which are evidenced not only by the extension of nominative case to these arguments but also by the use of expletive constituents and passive constructions, the present study is unable to make any definite claims.

In view of the limited evidence in support of the semantic hypothesis and the clear counterevidence presented by the case of *liken*, the question arises whether a more economical design of the present study would have been sufficient in obtaining the same level of results. The option of a smaller pilot study in which the basic predictions of the hypothesis are checked against the initial evidence may have been preferable, but this was difficult to realise due to the amount of pre-processing of the data required by the lack of comprehensive lemmatised corpora. This lack necessitated not only the individual extraction and assessment of potential verbal tokens from orthographically and morphologically diverse texts but also a relatively labour-intensive process of interpreting and annotating the ME and EModE sources, which frequently required extensive recourse to the surrounding context as well as the additional investigation of potentially non-distinctive morphological forms of pronominal *ye/you* in the later periods. For similar reasons, the use of quotations from the *MED* as the primary database would have been problematic, since these quotations often lack sufficient context for the identification of the relevant participants. While some indications about the associations of different types of stimuli with the construction type could have been obtained from a more limited survey, the present study fulfils the desire expressed by several scholars for a full corpus study of selected experiencer verbs and contributes to an improved understanding of the lexical diversity of their diachronic development. The employed statistical methods, particularly that of logistic regression modelling, provide a way of quantifying the observed associations, and their application to a set of diachronic data by implementation of a control variable that represents the general odds of the construction types at different periods in time can be regarded

as successful. It is, of course, regrettable that the limited amount of data and the uneven distribution of the variable levels did not facilitate such an approach with any of the verbs that were investigated besides *liken*. On the other hand, the fact that the available data would be generally insufficient for regression modelling could not have been anticipated in advance of their extraction and annotation, and the diachronically skewed distribution of different semantic types of stimuli presents a finding in itself. For *liken*, the notion of a diachronic change in lexical properties, reflected in its use with different semantic types of stimuli, can be entertained, but the observable impact of the difference in genre of the *PPCME2* and the *PCEEC* suggests that appropriate contexts for the occurrence of human stimuli involved in the literal experience of emotion may be simply absent in the ME corpus. The inclusion of the other corpus was necessary from a purely quantitative point of view and even advised by the assumed propensity of the correspondence to reflect linguistic changes in the spoken language with relative fidelity. The resulting break in genre that occurs in the data during the transition from the ME to the EModE period is otherwise undesirable, since it constitutes an additional dimension of complexity that makes the evaluation of already limited data even more difficult. In addition, the included correspondence betrays a much more formal register than expected, which only appears to be balanced by more private correspondence in the latter half of the corpus, and which results in the exceptional frequency of *liken* and *plesen* as markers of politeness. Without the inclusion of the *PCEEC*, the incidence of these verbs would probably be much closer to that of the other verbs and thus equally difficult to interpret in terms of the semantic hypothesis. A potential future endeavour would therefore be to complement or substitute the data from the *PCEEC* with those of the *Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME)*, which is more varied in terms genre and thus provides the opportunity of investigating the extent to which the findings of the present study can be replicated independently of the evidence of the correspondence. Regardless of whether or not the texts in that corpus would yield a sufficient number of observations to perform a quantitative analysis, it seems unlikely, however, that the principal results of the present study would change.

Bibliography

Primary Sources

- Alex. & Dind.* Skeat, Walter ed. (1878). *Alexander and Dindimus: Or, The Letters of Alexander to Dindimus, King of the Brahmins, with the Replies of Dindimus*. London: Trübner. (cited by line)
- Alfred, *Bede* Miller, Thomas ed. (1890–1898). *The Old English Version of Bede's Ecclesiastical History of the English People*. London: Trübner. (cited by book and chapter)
- Alfred, *Boeth.* Sedgefield, Walter ed. (1899). *King Alfred's Old English Version of Boethius De Consolatione Philosophiae*. Oxford: Clarendon Press. (cited by page number and line)
- Alfred, *C. P.* Sweet, Henry ed. (1871). *King Alfred's West-Saxon Version of Gregory's Pastoral Care, with an English Translation, the Latin Text, Notes, and an Introduction*. London: Trübner. (cited by page number and line)
- Alfred, *Oros.* Sweet, Henry ed. (1883). *King Alfred's Orosius*. London: Trübner. (cited by page number and line)
- Ælfric, *De Vet. Test.* Grein, Chr. ed. (1872). *Bibliothek der angelsächsischen Prosa*. Kassel, Göttingen: Wigand. (Text cited by page number and line)
- Ælfric, *Gen.* Grein, Chr. ed. (1872). *Bibliothek der angelsächsischen Prosa*. Kassel, Göttingen: Wigand. (Text cited by chapter and line)
- Ælfric, *Hom.* Skeat, Walter W. ed. (1881–1900). *Aelfric's Lives of Saints*. London: Trübner. (cited by volume, page number and line)

- Ælfric, Hom. G.* Godden, Malcolm ed. (1979). *Ælfric's Catholic Homilies: The Second Series*. London: Oxford University Press. (cited by page number and line)
- Ælfric, Hom. Th.* Thorpe, Benjamin ed. (1844–1846). *The Homilies of the Anglo-Saxon Church, The First Part, Containing the Sermones Catholici, or Homilies of Aelfric in the Original Anglo-Saxon*. London: Ælfric Society. (cited by volume, page number and line)
- Ælfric, Hom. P.* Pope, John ed. (1967). *Homilies of Ælfric: A Supplementary Collection*. London: Oxford University Press. (cited by homily and line number)
- Beow.* Heaney, Seamus ed. (2007). *Beowulf: A New Verse Translation*. London: Faber and Faber. (cited by line)
- Blickl. Homl.* Morris, Richard ed. (1880). *The Blickling Homilies of the Tenth Century*. London: Trübner. (cited by page number and line)
- Chaucer, Cant. T.* Skeat, Walter ed. (1894). *The Complete Works of Geoffrey Chaucer. The Canterbury Tales. Text*. Oxford: Clarendon Press. (cited by group and line)
- Chaucer, Poems* Skeat, Walter ed. (1899). *The Complete Works of Geoffrey Chaucer. Romaunt of the Rose. Minor Poems*. (cited by text and line)
- Chaucer, L. G. W.* Skeat, Walter ed. (1889). *Chaucer. The Legend of Good Women*. Oxford: Clarendon Press. (cited by line)
- Chron. Pl.* Plummer, Charles ed. (1892). *Two of the Saxon Chronicles Parallel, with Supplementary Extracts from the Others*. Oxford: Clarendon Press. (cited by page number)

- Curs. M.* Morris, Richard ed. (1874–1893). *Cursor Mundi: The Cursur o the world*. London: Trübner. (cited by line)
- Destr. of T.* Panton, G. and David Donaldson eds. (1869). *The “Gest hystoriale” of the Destruction of Troy: An Alliterative Romance tr. from Guido de Colonna’s “Hystoria troiana.”* London: Trübner. (cited by line)
- Exon. Th.* Thorpe, Benjamin ed. (1842). *Codex Exoniensis A Colection of Anglo-Saxon Poetry, from a Manuscript in the Library of the Dean and Chapter of Exeter*. London: Society of Antiquaries. (cited by page number and line)
- Genesis* Krapp, George Philip ed. (1931). *The Junius Manuscript*. New York: Columbia. (Text cited by line)
- Langland, Piers Plow.* Skeat, Walter ed. (1886). *The Vision of William Concerning Piers Ploughman. Vol. 1 – Text*. Oxford: Clarendon Press. (cited by passage and line)
- Marh.* Mack, Frances ed. (1934). *Seinte Marherete þe Meiden ant Martyr: Re-edited from Ms. Bodley 34, Oxford, and Ms. Royal 17A XXVii, British Museum*. London: Early English University Press. (cited after MED)
- More, Wks. 1557* Campbell, W. ed. (1931). *The English Works of Sir Thomas More. Reproduced in Facsimile from William Rastell’s Edition of 1557*. London: Eyre and Spottiswoode. (cited after Visser 1963)
- OE Gosp.* Thorpe, Benjamin ed. (1842). *The Anglo-Saxon Version of the Holy Gospels*. London: Rivington and Parker. (cited by text, chapter and line)

- OE Gosp. Mk. (Li)* Waring, George ed. (1861). *The Lindisfarne and Rushworth Gospels. Part II.* Durham: Andrews. (cited by chapter and line)
- Pol. Songs* Bøddeker, H. (1878). *Altenglische Dichtungen des Ms. Harl. 2253.* Berlin: Weidmannsche Buchhandlung. (cited by page number)
- Ps. Spl.* Spelman, J. ed. (1640). *Psalterium Davidis Latino-Saxonicum vetus.* London: Wallia. (cited by psalm and verse)
- Psalm Th.* Thorpe, Benjamin ed. (1835). *Libri Psalmorum Versio Antiqua Latina: Cum Paraphrasi Anglo-Saxonica, Partim Soluta Oratione, Partim Metrice Composita.* Oxford: Academic Press. (cited by psalm and verse)
- Shrn.* Cockayne, Thomas Oswald ed. (1864–1870). *The Shrine: A Collection of Occasional Papers on Dry Subjects.* London: Williams and Norgate. (cited by page number and line)
- Shute, Archit.* Weaver, Lawrence ed. (1912). *The First & Chief Groundes of Architecture. By John Shute, Paynter and Archytecte. First Printed in 1563. A Facsimile of the First Edition with an Introduction by Lawrence Weaver, F. S. A., Hon. A. R. I. B. A.* London: Country Life Ltd. (cited by folio)
- Will. of Pal.* Skeat, Walter ed. (1867). *The Romance of William of Palerne.* Trübner: London. (cited by line)
- Wulfila* Streitberg, Wilhelm ed. (1900). *Die gotische Bibel.* Heidelberg: Winter. (cited by text, chapter and line)
- Wulfstan, Hom.* Napier, A. S. ed. (1883). *Wulfstan. Sammlung der ihm zugeschriebenen Homilien.* Berlin: Weidmannsche Buchhandlung. (cited by page number and line)

Secondary Sources

Allen, Cynthia L.

- (1997) The Development of an “Impersonal” Verb in Middle English: The Case of *Behoove*. In: *Studies in Middle English Linguistics*. Ed. Jacek Fisiak. Berlin: de Gruyter. 1–21.
- (1995) *Case Marking and Reanalysis. Grammatical Relations from Old to Early Modern English*. Oxford: Clarendon Press.
- (1986a) Reconsidering the History of *Like*. *Journal of Linguistics* 22: 375–409.
- (1986b) Dummy Subjects and the Verb-Second Target in Old English. *English Studies* 67/6: 465–70.

Andersen, Henning

- (2001a) Introduction. In: *Actualization*. Ed. Henning Andersen. Amsterdam, Philadelphia: John Benjamins. 1–20.
- (2001b) Markedness and the theory of linguistic change. In: *Actualization*. Ed. Henning Andersen. Amsterdam, Philadelphia: John Benjamins. 21–57.
- (2001c) Actualization and the (uni)directionality of change. In: *Actualization*. Ed. Henning Andersen. Amsterdam, Philadelphia: John Benjamins. 225–248.
- (1973) Abductive and Deductive Change. *Language* 49/4: 765–793.

Andersen, Henning ed.

- (2001) *Actualization*. Amsterdam, Philadelphia: John Benjamins

Anderson, John M.

- (1988) The Type of Old English Impersonals. In: *Edinburgh Studies in the English Language*. Eds. John M. Anderson and Norman MacLeod. Edinburgh: John Donald Publishers Ltd. 1–32.
- (1986) A Note on Old English Impersonals. *Journal of Linguistics* 22: 167–77.
- (1984) The Natural History of Dative Sentences. In: *English Historical Linguistics: Studies in Development*. Eds. N. F. Blake and Charles Jones. Sheffield: The Centre for English Cultural Tradition and Language. 241–278.
- (1979) *On Being without a Subject*. Bloomington: Indiana University Linguistics Club.

Barðdal, Jóhanna et al. eds.

- (2015) *Diachronic Construction Grammar*. Amsterdam, Philadelphia: John Benjamins.

Barðdal, Jóhanna and Spike Gildea

- (2015) Diachronic Construction Grammar: Epistemological context, basic assumptions and historical implications. In: *Diachronic Construction Grammar*. Eds. Jóhanna Barðdal et al. Amsterdam, Philadelphia: John Benjamins. 1–49.

Barðdal, Jóhanna and Tomas Smitherman

- (2013) The Quest for Cognates: A Reconstruction of Oblique Subject Constructions in Proto-Indo-European. *Language Dynamics and Change* 3: 1–42.

Bauer, Brigitte

- (1998) Impersonal Verbs in Italic. Their Development from an Indo-European Perspective. *JIES* 26: 91–120.

Bellosta von Colbe, Valeriano

- (2004) Das indirekte Objekt als syntaktisches Argument ohne Makrorolle. In: *Semantische Rollen*. Eds. Rolf Kailuweit and Martin Hummel. Tübingen: Narr. 184–205.

Bergs, Alexander and Dieter Stein

- (2001) The Role of Markedness in the Actuation and Actualization of Linguistic Change. In: *Actualization*. Ed. Henning Andersen. Amsterdam, Philadelphia: John Benjamins. 79–93.

Blake, N. F. and Charles Jones eds.

- (1984) *English Historical Linguistics: Studies in Development*. Sheffield: The Centre for English Cultural Tradition and Language.

Blake, Frank R.

- (1930) A Semantic Analysis of Case. In: *Curme Volume of Linguistic Studies*. Ed. T. Hatfield, Werner Leopold and A. J. Friedrich Zieglschmid. Baltimore: Linguistic Society of America. 34–49.

Bosworth, Joseph and T. Northcote Toller

- (1898) *An Anglo-Saxon Dictionary*. Oxford: Clarendon Press.

Bubenik, Vit

- (2001) On the Actualization of the Passive-to-Ergative Shift in Pre-Islamic India. In: *Actualization*. Ed. Henning Andersen. Amsterdam, Philadelphia: John Benjamins. 95–118.

Busse, Ulrich

- (2001) The Use of Address Pronouns in Shakespeare's Plays and Sonnets. In: *Actualization*. Ed. Henning Andersen. Amsterdam, Philadelphia: John Benjamins. 119–142.

Bybee, Joan

- (2007) *Frequency of Use and the Organization of Language*. Oxford: University Press.
(2006) From Usage to Grammar: The Mind's Response to Repetition. *Language* 82/4: 711–733.

Brandenstein, Wilhelm

- (1928) Das Problem der Impersonalien. *IF* 46/1: 1–26.

Bródy, Mihály

- (1989) Old English Impersonals and the Theory of Grammar. *UCL Working Papers in Linguistics* 1: 262–294.

Brugmann, Karl

- (1925). *Die Syntax des einfachen Satzes im Indogermanischen*. Berlin: de Gruyter.
(1917) *Der Ursprung des Scheinsubjekts "es" in den germanischen und den romanischen Sprachen*. Leipzig: Teubner.

Charzyńska-Wójcik, Magdalena

- (2013) Feelings and Form in King Alfred's Psalter. *Language and Literary Studies of Warsaw* 3: 129–152.

Chung, Sandra

- (1977) On the Gradual Nature of Syntactic Change. In: *Mechanisms of Syntactic Change*. Ed. Charles Li. Austin, Tex., London: University of Texas Press. 3–55.

Cole, P., W. Harbert, G. Hermon and S. N. Sridhar

- (1980) The Acquisition of Subjecthood. *Language* 56/4: 719–743.
(1978) On the Acquisition of Subjecthood. *Studies in the Linguistic Sciences* 8: 42–71.

Comrie, Bernard

- (1993) Argument Structure. In: *Syntax. HSK 9.1*. Eds. Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld and Theo Vennemann. Berlin, New York: de Gruyter. 905–914.

Croft, William

- (2006) Evolutionary models and functional-typological theories of language change. In: *Handbook of the History of English*. Eds. Ans Kernenade, Bettelou Los. Oxford: Blackwell. 68–91.
(2000) *Explaining Language Change: An Evolutionary Approach*. Harlow: Longman.

Dąbrowska, Ewa

- (1994) Dative and nominative experiencers: two folk theories of the mind. *Linguistics* 32: 1029–1054.

Degaetano-Ortlieb, Stefania and Elke Teich

- (2018) Using Relative Entropy for Detection and Analysis of Periods of Diachronic Linguistic Change. In: *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Eds. Beatrice Alex, Stefania Degaetano-Ortlieb, Anna Feldman, Anna Kazantseva Nils Reiter and Stan Szpakowicz, Stan. Santa Fe, NM: Association for Computational Linguistics. 22–33.

Delbrück, Berthold

- (1900) *Vergleichende Syntax der indogermanischen Sprachen. Dritter Theil*. Straßburg: Trübner.

Denison, David

- (2008) Patterns and Productivity. In: *Studies in the history of the English language IV: Empirical and analytical advances in the study of English language change*. Eds. S. M. Fitzmaurice and D. Minkova. Berlin, New York: de Gruyter. 207–230.
- (1993) *English Historical Syntax: Verbal Constructions*. London, New York: Longman.
- (1990) The Old English Impersonals Revived. In: *Papers of the 5th International Conference on English Historical Linguistics*. Eds. Sylvia Adamson, Vivien A. Law, Nigel Vincent and Susan Wright. Amsterdam: John Benjamins. 111–141.
- (1989) Auxiliary + Impersonal in Old English. *Folia Linguistica Historica* 9/1: 139–166.

Denison, David et al. eds.

- (2012) *Analysing Older English*. Cambridge: University Press.

Diessel, Holger

- (2017) Usage-Based Linguistics. In: *Oxford Research Encyclopedia of Linguistics*. Ed. Mark Aronoff. New York: Oxford University Press.

Dowty, David

- (1991) Thematic Proto-Roles and Argument Selection. *Language* 67/3: 547-619.

Elmer, Willy

- (1981) *Diachronic Grammar: The History of Old and Middle English Subjectless Constructions*. Tübingen: Max Niemeyer Verlag.

Faarlund, Jan Terje

- (1990) *Syntactic Change. Toward a Theory of Syntactic Change*. Berlin, New York: de Gruyter.

Fillmore, Charles J.

- (1982) Frame Semantics. In: *Linguistics in the Morning Calm. Selected Papers from SICOL-1981*. Ed. The Linguistic Society of Korea. Seoul: Hanshin.
- (1977) The Case for Case Reopened. In: *Grammatical Relations*. Eds. Peter Cole and Jerrold M. Sadock. New York: Academic Press. 59–82.

- (1968) The Case for Case. In: *Universals in Linguistic Theory*. Eds. Emmon Bach and Robert T. Harms. New York: Holt, Rinehart and Winston. 1–88.

Fischer, Olga

- (2007) *Morphosyntactic Change: Functional and Formal Perspectives*. Oxford: University Press.
- (1992) Syntax. In: *The Cambridge History of the English Language. Volume II: 1066–1476*. Ed. Norman Blake. Cambridge: University Press.

Fischer, Olga, Hendrik De Smet and Wim van der Wurff

- (2017) *A Brief History of English Syntax*. Cambridge: University Press.

Fischer, Olga, Ans van Kemenade, Willem Koopman and Wim van der Wurff

- (2000) *The Syntax of Early English*. Cambridge: University Press.

Fischer, Olga and Frederike van der Leek

- (1987) A Case for the Old English Impersonal. In: *Explanation and Linguistic Change*. Eds. Willem Koopman, Frederike van der Leek, Olga Fischer and Roger Eaton. Amsterdam, Philadelphia: John Benjamins. 79–121.
- (1983) The Demise of the Old English Impersonal Construction. *Journal of Linguistics* 19: 337–68.

Fischer, Susann

- (2013) Subjects, subject positions and word-order: Old Spanish vs. Old English. In: *Sprachwandelvergleich – Comparing Diachronies*. Eds. Jürg Fleischer and Horst J. Simon. Berlin, Boston: de Gruyter. 119–140.

Fisiak, Jacek

- (1968) *A Short Grammar of Middle English. Part One: Graphemics, Phonemics and Morphemics*. Warszawa: Państwowe wydawnictwo naukowe.

Fleischer, Jürg and Horst J. Simon

- (2013) Vom Wert sprachwandelvergleichenden Arbeitens. In: *Sprachwandelvergleich – Comparing Diachronies*. Eds. Jürg Fleischer and Horst J. Simon. Berlin, Boston: de Gruyter. 1–14.

Foley, William A. and Robert D. Van Valin Jr.

- (1984) *Functional Syntax and Universal Grammar*. Cambridge: University Press.

Friginal, Eric and Jack A. Hardy

- (2014) *Corpus-Based Sociolinguistics*. New York, London: Routledge.

van der Gaaf, W.

- (1904) *The Transition from the Impersonal to the Personal Construction in Middle English. Anglistische Forschungen 14*. Heidelberg: Carl Winter.

Gast, Volker and Natalia Levshina

- (2014) Motivating w(H)-Clefts in English and German: A Hypothesis-Driven Parallel Corpus Study. In: *Frequency, Forms and Functions of Cleft Constructions in Romance and Germanic. Contrastive, Corpus-Based Studies*. Ed. Anna-Maria De Cesare. Berlin: de Gruyter.

Goebel, Julius

- (1888) On the Impersonal Verbs. *Transactions of the American Philological Association (1869-1896)*. 19: 20–32.

Goldberg, Adele E.

- (1995) *A Construction Grammar Approach to Argument Structure*. Chicago, London: The University of Chicago Press.

Gries, Stefan Th.

- (²2017) *Quantitative Corpus Linguistics with R. A Practical Introduction*. New York, London: Routledge.
- (²2013) *Statistics for Linguistics with R. A Practical Introduction*. Berlin, Boston: de Gruyter.

(2008) Dispersions and adjusted frequencies in corpora. *IJCL* 13/4: 403–437.

Gries, Stefan Th. and Martin Hilpert

(2008) The Identification of Stages in Diachronic Data: Variability-Based Neighbour Clustering. *Corpora* 3/1: 59–81.

Gruber, Jeffrey S.

(1965) *Studies in Lexical Relations*. Doctoral dissertation, MIT, Cambridge MA.

Harris, Alice C. and Lyle Campbell

(1995) *Historical Syntax in Cross-Linguistic Perspective*. *Cambridge Studies in Linguistics* 74. Cambridge: University Press.

Harweg, Roland

(1969) Bemerkungen zum sogenannten Identitätspronomen *derselbe*. *Zeitschrift für Dialektologie und Linguistik*. 36/3: 269–303.

Hilpert, Martin and Stefan Th. Gries

(2016) Quantitative Approaches to Diachronic Corpus Linguistics. In: *The Cambridge Handbook of English Historical Linguistics*. Eds. Merja Kytö and Päivi Pahta. Cambridge: University Press. 36–53.

Hogg, Richard

(2006) Old English Dialectology. In: *Handbook of the History of English*. Eds. Ans Kernenade, Bettelou Los. Oxford: Blackwell. 395–416.

Holthausen, F.

(1934) *Altenglisches etymologisches Wörterbuch*. Heidelberg: Carl Winter.

Hopper, Paul J. and Sandra A. Thompson

(1980) Transitivity in Grammar and Discourse. *Language* 56/2: 251–299.

Hopper, Paul J. and Elizabeth C. Traugott

(²2003) *Grammaticalization*. Cambridge: University Press.

Horobin, Simon and Jeremy Smith

- (2002) *An Introduction to Middle English*. Edinburgh: University Press.

Hubbard, Frank G.

- (1918) Beowulf 1598, 1996, 2026; Uses of the Impersonal Verb *Geweorþan*. *The Journal of English and Germanic Philology*. 17/1: 119–124.

Jackendoff, Ray

- (1987) The Status of Thematic Relations in Linguistic Theory. *Linguistic Inquiry* 18/3: 369–411.

Jacob, Daniel

- (2004) Generalisierte Semantische Rollen und freie Linkingprinzipien. In: *Semantische Rollen*. Eds. Rolf Kailuweit and Martin Hummel. Tübingen: Narr. 104–117.

Jespersen, Otto

- (1949) *A Modern English Grammar on Historical Principles. Part VII. Syntax. Completed and Edited by Niels Haislund*. Copenhagen: Ejnar Munksgaard.
- (1927) *A Modern English Grammar on Historical Principles. Part III. Syntax (Second Volume)*. Copenhagen: Ejnar Munksgaard.
- (1894) *Progress in Language*. London: Sonnenschein.

Kailuweit, Rolf

- (2004) Protorollen und Makrorollen. In: *Semantische Rollen*. Eds. Rolf Kailuweit and Martin Hummel. Tübingen: Narr. 83–103.
- (2002) Pragmatische Gründe syntaktischer Flexibilität? "psych verbs" im Spanischen und in anderen romanischen Sprachen. In: *Gramática y Pragmática del Español. Actas de la sección Grammatik und Pragmatik im Spanischen del XIII. Deutscher Hispanistentag, Leipzig (8.–11. III. 2001)*. Eds. Alberto Gil and Christian Schmitt. 57–90.

Kailuweit, Rolf, Carmen Kelling and Judith Meinschaefer

- (2003) Argumentvererbung und Argumentrealisierung bei Nominalisierungen von psychologischen Verben. In: *Syntaxtheorien. Modelle, Methoden, Motive*. Eds. Eilsabeth Stark and Ulrich Wandruszka. 71–101.

Kailuweit, Rolf and Martin Hummel

- (2004) Einführung: Semantische Rollen in der Diskussion. In: *Semantische Rollen*. Eds. Rolf Kailuweit and Martin Hummel. IX–XXIX.

Kailuweit, Rolf and Martin Hummel eds.

- (2004) *Semantische Rollen*. Tübingen: Narr.

Keenan, Edward L.

- (1976) Towards a Universal Definition of Subject. In: *Subject and Topic*. Ed. Charles N. Li. New York: Academic Press. 303–333.

Kemenade, Ans and Bettelou Los eds.

- (2006) *Handbook of the History of English*. Oxford: Blackwell.

Kim, Hyeree

- (1999) A lexical approach to the history of the quasi-impersonal subject ‘it’. *English Studies* 80/4: 318–342.

Klimov, Georgij

- (1974) On the Character of Languages of Active Typology. *Linguistics* 131: 11–25.

Komen, Erwin R.

- (2017) Beyond Counting Syntactic Hits. In: *CLARIN in the Low Countries*. Eds. J. Odijk and A. van Hessen. London: Ubiquity Press. 259–268.

Krzyszpień, Jerzy

- (1990) *A Diachronic Analysis of English Impersonal Constructions with an Experiencer*. Kraków: Nakład Uniwersytetu Jagiellońskiego.

Labov, William

- (1972) The Recent History of Some Dialect Markers on the Island of Martha's Vineyard, Mass. In: *Studies in Linguistics in Honor of Raven I. McDavid Jr.*. Ed. Lawrence M. Davis. University of Alabama: University of Alabama Press. 81–121.
- (1966) *The Social Stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.

Lagerquist, Linnea M.

- (1985) The Impersonal Verb in Context: Old English. In: *Papers from the 4th International Conference on English Historical Linguistics, Amsterdam, April 10–13, 1985*. Eds. Roger Eaton, Olga Fischer, Willem F. Koopman and Frederike van der Leek. 123–136.

Langacker, Ronald W.

- (1977) Syntactic Reanalysis. In: *Mechanisms of Syntactic Change*. Ed. Charles Li. Austin, Tex., London: University of Texas Press. 57–139.

Levshina, Natalia

- (2015) *How to do Linguistics with R. Data Exploration and Statistical Analysis*. Amsterdam, Philadelphia: John Benjamins.

Lewis, Robert E. and Mary Jane Wililams

- (²2007) *Middle English Dictionary: Plan and Bibliography*. Ann Arbor: University of Michigan Press.

Lightfoot, David W.

- (1999) *The Development of Language: Acquisition, Change, and Evolution*. Malden, Mass.: Blackwell.
- (1991) How to set Parameters: Arguments from Language Change. *Language* 70/1: 141–152.
- (1981) Explaining syntactic change. In: *Explanation in Linguistics*. Eds. Norbert Hornstein and David Lightfoot. London, New York: Longman. 209–240.
- (1979) *Principles of Diachronic Syntax*. Cambridge: University Press.

Lijffijt, Jeffrey and Stefan Th. Gries

- (2012) Correction to Stefan Th. Gries' "Dispersions and adjusted frequencies in corpora". *IJCL* 17/1: 147–149.

Lindemann, John

- (1970) *Old English Preverbal 'Ge'. Its Meaning*. Charlottesville: University of Virginia.

Los, Bettelou

- (2009) The Consequences of the Loss of Verb-Second in English: Information Structure and Syntax in Interaction. *English Language and Linguistics* 13/1: 97–125.

Loureiro-Porto, Lucía

- (2010) A Review of Early English Impersonals: Evidence from Necessity Verbs, *English Studies* 91/6: 674–699.
- (2009) *The Semantic Predecessors of Need in the History of English (c750–1710)*. Malden, MA: Blackwell.

Maling, Joan and Sigríður Sigurjónsdóttir

- (2002) The 'new impersonal' construction in Icelandic. *JCGL* 5: 97–142.

Marcus, Mitchell P., Beatrice Santorini and Mary Ann Marcinkiewicz

- (1993) Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics* 19/2: 313–330.

Markus, Manfred

- (1990) *Mittelenglisches Studienbuch*. Tübingen: Francke.

McCawley, Noriko A.

- (1976) From OE/ME 'Impersonal' to 'Personal' Constructions: What is a 'Subject-less' S? In: *Papers from the Parasession on Diachronic Syntax, Chicago Linguistic Society*. Eds. Sandfor B. Steever, Alice C. Walker and Salikoko S. Mufwene. Chicago: Linguistic Society. 192–204.

McMahon, A.

- (1994) *Understanding Language Change*. Cambridge: University Press.

McIntosh, Angus, Michael L. Samuels and Michael Benskin

- (1986) *A Linguistic Atlas of Late Mediaeval English*. 4 vols. Aberdeen: University Press.

Méndez-Naya, Belén and María José López-Couso

- (1997) What is Really Meant by Impersonal? On Impersonal and Related Terms. *Atlantis* 19/2: 185–92.

Milroy, Lesley

- (1980) *Language and Social Networks*. Oxford: Blackwell.

Milroy, James and Lesley Milroy

- (1978) Belfast: Change and Variation in an Urban Vernacular. In: *Sociolinguistic Patterns in British English*. Ed. Peter Trudgill. 19–36.

Mitchell, Bruce

- (1990) *A Critical Bibliography of Old English Syntax: To the End of 1984, Including Addenda and Corrigenda to "Old English Syntax"*. Oxford: Blackwell.
- (1985) *Old English Syntax: Concord, The Parts of Speech, and the Sentence*. Oxford: Clarendon Press.

Mitchell, Bruce and Fred C. Robinson

- (⁷2007) *A Guide to Old English*. Malden, MA: Blackwell.

Mithun, Marianne

- (2001) Actualization Patterns in Grammaticalization: From Clause to Locative Morphology in Northern Iroquoian. In: *Actualization*. Ed. Henning Andersen. Amsterdam, Philadelphia: John Benjamins. 143–168.

Miura, Ayumi

- (2015) *Middle English Verbs of Emotion and Impersonal Constructions. Verb Meaning and Syntax in Diachrony*. Oxford: University Press.

Möhlig-Falke, Ruth

- (2016) Using the Dictionary of Old English Corpus for Linguistic Analyses: A Basic Classification of the Textual Sources. *Neuphilologische Mitteilungen* 116/2: 395–420.
- (2012) *The Early English Impersonal Construction. An Analysis of Verbal and Constructional Meaning*. Oxford: University Press.

Mossé, Fernand

- (1952) *A Handbook of Middle English*. Baltimore: Hopkins.

Nevalainen, Terttu

- (1996) Introduction. In: *Sociolinguistics and Language History. Studies Based on the Corpus of Early English Correspondence*. Eds. Terttu Nevalainen and Helena Raumolin-Brunberg. Amsterdam, Atlanta, GA: Rodopi. 3–9.
- (1996) Social Stratification. In: *Sociolinguistics and Language History. Studies Based on the Corpus of Early English Correspondence*. Eds. Terttu Nevalainen and Helena Raumolin-Brunberg. Amsterdam, Atlanta, GA: Rodopi. 57–76.
- (1996) Gender Difference. In: *Sociolinguistics and Language History. Studies Based on the Corpus of Early English Correspondence*. Eds. Terttu Nevalainen and Helena Raumolin-Brunberg. Amsterdam, Atlanta, GA: Rodopi. 77–91.

Nevalainen, Terttu and Helena Raumolin-Brunberg

- (1996) The Corpus of Early English Correspondence. In: *Sociolinguistics and Language History. Studies Based on the Corpus of Early English Correspondence*. Eds. Terttu Nevalainen and Helena Raumolin-Brunberg. Amsterdam, Atlanta, GA: Rodopi. 39–54.

Nevalainen, Terttu and Helena Raumolin-Brunberg eds.

- (1996) *Sociolinguistics and Language History. Studies Based on the Corpus of Early English Correspondence*. Amsterdam, Atlanta, GA: Rodopi.

Nichols, Johanna

- (1984) Functional Theories of Grammar. *Annual Review of Anthropology* 13: 97–117.

Ogura, Michiko

- (2013) *Words and Expressions of Emotion in Medieval English*. Frankfurt: Peter Lang.
- (1990) What Has Happened to the Impersonal Constructions? *Neophilologische Mitteilungen* 91/1: 31–55.
- (1986) *Old English 'Impersonal' Verbs and Expressions*. *Anglistica* 24. Copenhagen: Rosenkilde and Bagger.

Palander-Collin, Minna

- (1997) A Medieval Case of Grammaticalization, *methinks*. In: *Grammaticalization at Work. Studies of Long-Term Developments in English*. Eds. Matti Rissanen, Merja Kytö and Kirsi Heikkonen. Berlin, New York: de Gruyter. 371–403.
- (1996) The Rise and Fall of METHINKS. In: *Sociolinguistics and Language History. Studies Based on the Corpus of Early English Correspondence*. Eds. Terttu Nevalainen and Helena Raumolin-Brunberg. Amsterdam, Atlanta, GA: Rodopi. 131–149.

Percillier, Michael

- (2018) A Toolkit for Lemmatizing, Analysing, and Visualising Middle English Data. In: *CRH-2: proceedings of the Second Workshop on Corpus-based Research in the Humanities 25-26 January 2018 Vienna, Austria*. Ed. A. U. Frank. Wien: Dept. of Geoinformation, TU Wien. 153–160.
- (2016) Verb Lemmatization and Semantic Verb Classes in a Middle English Corpus. In: *Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016) Bochum, Germany September 19-21*. Eds. Stefanie Dipper, Friedrich Neubarth and Heike Zinsmeister. Bochum: Ruhr-Universität. 209–214.

Pishwa, Hanna

- (1999) The Case of the “Impersonal” Construction in Old English. *Folia Linguistica Historica* 20/1–2: 129–151.

Pocheptsov, George G.

- (1997) Quasi-Impersonal Verbs in Old and Middle English. In: *Studies in Middle English Linguistics*. Ed. Jacek Fisiak. Berlin: de Gruyter. 469–488.

Primus, Beatrice

- (2004) Protorollen und Verbtyp: Kasusvariation bei psychischen Verben. In: *Semantische Rollen*. Eds. Rolf Kailuweit and Martin Hummel. Tübingen: Narr. 377–401.
- (1999) *Cases and Thematic Roles. Ergative, Accusative and Active*. Tübingen: Narr.
- (1993) Syntactic Relations. In: *Syntax. HSK 9.1*. Eds. Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld and Theo Vennemann. Berlin, New York: de Gruyter. 686–705.

Raumolin-Brunberg, Helena

- (1996) Historical Sociolinguistics. In: *Sociolinguistics and Language History. Studies Based on the Corpus of Early English Correspondence*. Eds. Terttu Nevalainen and Helena Raumolin-Brunberg. Amsterdam, Atlanta, GA: Rodopi. 11–37.
- (1996) Forms of Address in Early English Correspondence. In: *Sociolinguistics and Language History. Studies Based on the Corpus of Early English Correspondence*. Eds. Terttu Nevalainen and Helena Raumolin-Brunberg. Amsterdam, Atlanta, GA: Rodopi. 167–181.

Rissanen, Matti

- (2000) The World of English Historical Corpora. From Cædmon to the Computer Age. *Journal of English Linguistics* 28/1: 7–20.
- (1999) Syntax. In: *The Cambridge History of the English Language. Volume III: 1476–1776*. Ed. Roger Lass. Cambridge: University Press.
- (1987) Review: *English Historical Linguistics: Studies in Development*. Eds. N. F. Blake and Charles Jones. Sheffield: University of Sheffield. 1984. In: *JEngL* 20/1: 150–156.

Rissanen, Matti, Merja Kytö and Kirsi Heikkonen eds.

- (1997) *Grammaticalization at Work. Studies of Long-Term Developments in English*. Berlin, New York: de Gruyter.

Roberts, Ian

- (2007) *Diachronic Syntax*. Oxford: University Press.

Ross, John Robert

- (1972) The Category Squish: Endstation Hauptwort. In *Papers from the Eighth Regional Meeting. Chicago Linguistic Society. April 14–16, 1972*. Eds. Paul M. Peranteau, Judith N. Levi and Gloria C. Phares. Chicago: Linguistic Society.

von Seeffranz-Montag, Ariane

- (1984) ‘Subjectless’ Constructions and Syntactic Change. In: *Historical Syntax. Trends in Linguistics/Studies and Monographs 23*. Ed. Jacek Fisiak. Paris, The Hague: Mouton de Gruyter. 521–553.
- (1983) *Syntaktische Funktionen und Wortstellungsveränderung. Die Entwicklung „subjektloser“ Konstruktionen in einigen Sprachen*. München: Fink.

De Smet, Hendrik

- (2012) The Course of Actualization. *Language* 88/3: 601–633.

Smith, John Charles

- (2001) Markedness, Functionality, and Perseveration in the Actualization of a Morphosyntactic Change. In: *Actualization*. Ed. Henning Andersen. Amsterdam, Philadelphia: John Benjamins. 203–223.

Taylor, Ann, Mitchell Marcus and Beatrice Santorini

- (2003) The Penn Treebank: An Overview. In: *Treebanks: Building and Using Parsed Corpora*. Ed. Anne Abeillé. Dordrecht, Boston, London: Kluwer Academic Publishers. 5–22.

Timberlake, Alan

- (1977) Reanalysis and Actualization in Syntactic Change. In: *Mechanisms of Syntactic Change*. Ed. Charles Li. Austin, Tex., London: University of Texas Press. 141–177.

Toller, T. Northcote

- (1921) *An Anglo-Saxon Dictionary. Supplement.* Oxford: Clarendon Press.

Traugott, Elizabeth C.

- (2015) Toward a coherent account of grammatical constructionalization. In: *Diachronic Construction Grammar*. Eds. Jóhanna Barðdal et al. Amsterdam, Philadelphia: John Benjamins. 51–79.
- (1992) Syntax. In: *The Cambridge History of the English Language. Volume I: The Origins to 1066*. Ed. Richard Hogg. Cambridge: University Press.

Trips, Carola and Achim Stein

- (2019) Contact-Induced Changes in the Argument Structure of Middle English Verbs on the Model of Old French. *Journal of Language Contact* 12/1: 232–267.

Trousdale, Graeme

- (2008) Words and Constructions in Grammaticalization: The End of the English Impersonal Construction. *Studies in the history of the English language IV: Empirical and analytical advances in the study of English language change*. Eds. S. M. Fitzmaurice and D. Minkova. Berlin, New York: de Gruyter. 301–326.

Van Valin, Robert D. Jr.

- (2004) Semantic Macroroles in Role and Reference Grammar. In: *Semantische Rollen*. Eds. Rolf Kailuweit and Martin Hummel. Tübingen: Narr. 62–82.

Van Valin, Robert D. Jr. and Randy J. La Polla

- (1997) *Syntax. Structure, Meaning and Function*. Cambridge: University Press.

Visser, Fredericus T.

- (1963) *An Historical Syntax of the English Language. Part One. Syntactical Units with One Verb*. Leiden: Brill.

Wahlén, Nils Johann

- (1925) *The Old English impersonalia. Part I. Impersonal Expressions Containing Verbs of Material Import in the Active Voice*. Gothenburg: Elanders Boktryckeri.

Waltz, Heidi

(2000) (Im)personal Verbs of Emotion in Germanic. *IF* 105: 284–303.

Williams, E.

(1984) Grammatical Relations. *Linguistic Inquiry* 15: 639–73.

Wotjak, Gerd

(2004) Partizipantenrollen in Sachverhaltenswissensrepräsentationen und als semantisch-funktionale Argumentbestimmungen in Verbbedeutungen. In: *Semantische Rollen*. Eds. Rolf Kailuweit and Martin Hummel. Tübingen: Narr. 3–36.

van der Wurff, Wim

(1992) Another Old English Impersonal: Some Data. In: *Evidence for Old English: Material and Theoretical Bases for Reconstruction*. Ed. Fran Colman. Edinburgh: John Donald. 211–248.

Digital Resources

An Anglo-Saxon Dictionary Online.

<http://bosworth.ff.cuni.cz/>

(last updated 16 December 2013)

BASICS Toolkit. Release 4.

<http://bwcloud-108-183.bwcloud.uni-mannheim.de/>

(last updated September 2020)

Middle English Dictionary.

<https://quod.lib.umich.edu/m/med/>

(last revised in 2018)

Oxford English Dictionary.

<https://www.oed.com/>

(last updated in December 2020)

Parsed Corpus of Early English Correspondence. Eds. Terttu Nevalainen Helena Raumolin-Brunberg, Jukka Keränen et al. Oxford Text Archive.
<http://hdl.handle.net/20.500.12024/2510>
(released in 2006)

The Penn-Helsinki Parsed Corpus of Middle English (PPCME2). Eds. Anthony Kroch and Ann Taylor. Department of Linguistics, University of Pennsylvania. CD-ROM, second edition, release 2.
<http://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-2/>
(released in 2000)

Python Software Foundation. Python Language Reference.
<http://www.python.org/>
(version 3.7, released 27 June 2018)

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
<https://www.R-project.org/>
(version 3.6.1, released 5 July 2019)

Ehrenwörtliche Erklärung

Hiermit erkläre ich, dass mir die geltende Promotionsordnung der Philosophischen Fakultät der Friedrich-Schiller-Universität Jena bekannt ist und dass ich die vorgelegte Dissertation selbst angefertigt, keine Textabschnitte eines anderen Autors oder eigener Prüfungsarbeiten ohne Kennzeichnung übernommen und alle von mir benutzten Hilfsmittel und Quellen in meiner Arbeit angegeben habe.

Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskriptes haben mich keine anderen Personen unterstützt; die Hilfe eines Promotionsberaters habe ich nicht in Anspruch genommen und Dritte haben weder unmittelbar noch mittelbar geldwerte Leistungen von mir für die Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Des Weiteren erkläre ich, dass die Dissertation noch nicht als Prüfungsarbeit für eine wissenschaftliche Prüfung eingereicht habe; auch habe ich keine gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung bei einer anderen Hochschule als Dissertation eingereicht.

Statement on honour

I hereby declare that I am aware of the applicable doctoral regulations of the Faculty of Arts at the Friedrich Schiller University of Jena and that I have completed the submitted dissertation myself, that I have not included text sections from another author or my own examination theses without marking them and that I have indicated all resources and sources used by me in my dissertation.

No other individuals have assisted me in the selection and evaluation of the material or in the preparation of the manuscript; I have not enlisted the help of a doctoral consultant and third parties have not received any direct or indirect monetary benefits from me for work relating to the submitted dissertation.

Furthermore, I declare that I have not submitted this doctoral thesis as an academic examination paper elsewhere; nor have I submitted an identical or substantially similar thesis to another university as a doctoral dissertation.

Ort, Datum:

Unterschrift: