On Clausal Subjects and Extraposition in the History of English

On Clausal Subjects and Extraposition in the History of English

Rickard Ramhöj

UNIVERSITY OF
GOTHENBURG

Doctoral dissertation in English, University of Gothenburg
April 15, 2016

Dissertation edition, April 2016. Cover design: Liza Claesson
Printed by Reprocentralen Campusservice Lorensberg, University of Gothenburg, 2016
ISBN: 978-91-979921-7-6
http://hdl.handle.net/2077/41583

Distribution: Department of Languages and Literatures, University of Gothenburg, Box 200, SE-40530 Göteborg

TITLE: On Clausal Subjects and Extraposition in the History of English
LANGUAGE: English
AUTHOR: Rickard Ramhöj
DISTRIBUTION: Department of Languages and Literatures, University of Gothenburg, Box 200, SE-40530 Göteborg
ISBN: 978-91-979921-7-6
http://hdl.handle.net/2077/41583

Abstract

This study deals with disputed issues in the history of English concerning predicates that alternately take (i) a preposed clausal subject and (ii) a subject *it* in conjunction with a propositional subclause. Situated within the theoretical framework of Lexical Functional Grammar and based on present-day and historical corpora of English, the dissertation presents a number of claims with respect to the syntax and argument structure as well as the pragmatic and processing-related aspects of the relevant constructions. It is shown that, while all types of clauses can be analysed as morphosyntactic subjects in Early and Late Modern English, only infinitival clauses, and not *that*-clauses or *wh*-clauses, can be analysed as structural subjects. In Old and Middle English, the data is inconclusive as to the analysis of subclauses as subjects.

With respect to the co-occurrence of a subject *it* and a propositional subclause, two distinct constructions are recognised: (i) *it*+ADJ and (ii) *it*+COMP. *It*+ADJ has a thematic subject *it* in conjunction with an adjunct subclause, while *it*+COMP has a non-thematic subject *it* in conjunction with a complement subclause. *It*+ADJ is available at all stages of the history of English, while *it*+COMP seems to emerge in connection to the development of raising verbs.

Concerning pragmatic and processing-related aspects of the constructions, weight, complexity and information structure all have considerable effects on the choice of construction in both present-day and historical English. For the Present-day English data, it is shown that there is a cut-off point in the weight distribution of the constructions favouring one construction or the other. It is also shown that subclauses lacking an anaphoric relation to the previous discourse exclusively occur in the *it*+subclause construction, while subclauses expressing polar contrast exclusively occur in the preposed clausal subject construction.

KEYWORDS: clausal subjects, extraposition, History of English, argument structure, weight, complexity, information structure, Lexical-Functional Grammar, Lexical Mapping Theory

Acknowledgment

I wish to acknowledge the support of the Department of Languages and Literatures at the University of Gothenburg for my employment, which made it possible for me to pursue this project. Acknowledgement is furthermore warranted for the support of Kungliga och Hvitfeldtska stiftelsen for a scholarship covering the last six months of the project, and for the support of Adlerbertska Stipendiestiftelsen and Stiftelsen Paul och Marie Berghaus donationsfond, who generously provided funding for me to attend *The Sixth Graduate International Summer School in Cognitive Sciences and Semantics* in Riga, Latvia, and *The 20th International Lexical Functional Grammar Conference* in Tokyo, Japan. I am grateful to the participants and organizers of these events, as well as the participants and organizers of *The 9th New York - St. Petersburg Institute of Linguistics, Cognition and Culture* in St. Petersburg, Russia, *The 2013 Linguistic Institute* in Ann Arbor, Michigan, USA, *The Annual Meeting of the Linguistic Association of Great Britain* in Oxford, UK, *Understanding pro-drop. A synchronic and diachronic perspective* in Trento, Italy, and *The 9th Leiden Summer School in Languages and Linguistics* in Leiden, Holland.

As for particular individuals who have contributed to the project, I first want to mention my two supervisors Gunnar Bergh and Maia Andréasson. I would like to thank both of them for their guidance and support, as well as for reading through and commenting on countless badly written drafts. I would also like to thank Gerlouf Bouma, Henrik Rosenkvist, Hubert Cuyckens and Filippa Lindahl, who have read and commented on parts of the text at various stages of the process. Without their input, the text would have looked radically different, much for the worse.

I have also had some help with the analysis and judgement of sentences from languages other than English. I would like to thank Andreas Keränen and Erik Bohlin for their help with Latin, and Heike Havermeier and Michelle Waldispühl for their help with German.

Many colleagues at the University of Gothenburg have provided invaluable input at seminars, courses, reading groups, lunches, coffees and dinners. I want to thank all of my collegues, teachers and fellow classmates. Special thanks are due to 'the Luncheonists', who have provided crucial support and mental recreation during our daily luncheons.

Göteborg, March 2016

Rickard Ramhöj

Table of contents

List of Figures

List of Tables

List of Abbreviations

Grammaticality judgements

*	ungrammatical
%	grammatical for some speakers
?	of questionable grammaticality
#	grammatical, but pragmatically infelicitous

Syntactic categories

ADV	adverb
ADJ	adjective
C	complementizer
D	determiner
I	inflection
N	noun
P	phrase
V	verb
S	exocentric sentence

Stages of English

OE	Old English (450-1100)
ME	Middle English (1100-1500)
EME	Early Modern English (1500-1710
LME	Late Modern English (1710-1914)
PDE	Present-day English

Feature specifications

CASE	case
NOM	nominative
ACC	accusative
GEN	genitive
DAT	dative
NUM	number
SG	singular
PL	plural
PERS	person
1ST	first person
2ND	second person
3RD	third person
GEN	gender
MASC	masculine
FEM	feminine
NEU	neuter

Grammatical functions

ADJ	adjunct function
AF	argument function
COMP	sentential complement
GF	grammatical function
OBJ	object
OBJ_θ	thematically restricted object
OBL_θ	thematically restricted oblique
SUBJ	subject
UDF	unbounded dependency function
XCOMP	functionally controlled infinitive

Chapter 1

Introduction

1.1 Background

The present study deals historically with a much-discussed syntactic phenomenon in English, often referred to as *extraposition of a sentential subject* (e.g. Rosenbaum, 1967) or *it-extraposition* (e.g. Kaltenböck, 2005). It concerns the alternation between two different constructions: firstly, a configuration with a clausal (sometimes called sentential) subject, i.e. a finite or non-finite subordinate clause acting as a subject within a superordinate clause, and, secondly, a configuration with a subject *it* in conjunction with a propositional subclause. These two constructions are exemplified in (1) with sentences taken from the historical corpora used in the study. In this case, both sentences derive from the Early Modern English period.

(1) a. The preposed clausal subject construction
That a Solution of Silver does Dye Hair of a Black Colour, is a Known Experiment, ...
(BOYLECOL-E3-P2,150.80)

b. The *it*+subclause construction
It is sayde that I wuld haue saued the senators.
(BOETHCO-E1-P1,20.47)

In (1-a), we see an example of a subclause, *that a Solution of Silver does Dye Hair of a Black Colour*, occurring in a clause-initial position followed by the copula *is* and a complement, *a Known Experiment*. The term used for this kind of construction will be *preposed clausal subject*. It is referred to as preposed because the subclause is clause-initial, subclauses otherwise typically occurring in a clause-final position, and the term clausal subject

is used because there is evidence to suggest that the subordinate clause in this type of construction is to be analysed as a morphosyntactic subject (e.g. Huddleston & Pullum, 2002: 957).

In (1-b), the subordinate clause *that I would haue saued the senators* occurs in a clause-final position while the typical subject position, immediately preceding the finite verb, is occupied by the pronoun *it*. Descriptively, this construction will be referred to as the *it+subclause construction*.

The subject *it* in the *it*+subclause construction is more or less obligatory in Present-day English. This is to say that we do not find examples of predicates taking only a clausal argument, where the clausal argument occurs in a clause-final position. In earlier periods of English, this was not the case, and clause-final subclauses were not always accompanied by a pronoun *it* in subject position. An example from the Old English period of a clause-final subclause without any nominal subject constituent is given in (2).

(2) The null+subclause construction
Gregorius cwæð, on sumum timan gelamp, þæt sum man
Gregory said on some time happened that some man
forlet his eagena gesihðe.
lost his eyes' sight
'Gregory said that it happened at one time that some man lost his eyesight.'
(cogregdH,GD_1_[H]:10.77.18.761)

In this example, there is no subject pronoun *it* even though the subclause *þæt sum man forlet his eagena gesihðe* occurs in a clause-final position. This construction will be descriptively referred to as the *null+subclause construction*.

1.2 Statement of the problem

With respect to the sentences in (1) and (2), there are a number of disputed points in the literature. In this study, I attempt to give an answer to some of these issues based on material from historical and present-day corpora of English.

With respect to sentences such as the one in (1-a), the preposed clausal subject construction, the disputed points concern the syntactic and morphosyntactic properties of the subclause in the history of Engl-ish. Opinions differ as to the structural position of the subclause in sentences

such as (1-a), whether it occurs in the typical subject position of English, Spec,IP, or whether it occurs in a fronted position above. Depending on the framework assumed, opinions also differ as to the status of the subclause, whether it constitutes a subject or a fronted complement. With respect to the early stages of English, and in particular Old English, there is also a question whether subordinate clauses can function as subjects at all, and whether clause-final subclauses such as the one in (2), an example of the null+subclause construction, constitute subjects or complements.

With respect to the sentence in (1-b), the *it*+subclause construction, the main points of interest concern the status of the subject *it* as well as the status of the propositional subclause. Put concisely, the central questions are if the subject *it* is thematic or non-thematic, and if the subclause constitutes a complement or an adjunct.

Apart from questions concerning the syntax and argument structure of the sentences in (1) and (2), another disputed point in the literature concerns the ways in which weight, complexity and information structure influence the choice of construction, when both alternatives are possible. The questions here concern to what extent weight, complexity and information structure can account for the choice of construction, and, furthermore, what the relevant aspects of these factors are. Is there a particular tipping point in terms of relative weight, when one or the other construction is preferred? Is it givenness, contrast, or the activation of discourse referents that is relevant in the choice of construction?

1.3 Delimitation of object of investigation

As we have already seen, the subclauses in (1) and (2) constitute *that*-clauses. The alternation between the preposed clausal subject construction and the *it*+subclause construction can, however, also occur with other types of clauses. In (3), examples of the alternation with additional types of subclauses are given.

(3) a. Interrogative clauses
   (i) & [whether I ever get beyond the first] is doubtful.
   (AUSTEN-180X,176.350)
   (ii) it is doubtful [whether the King will condescend to what the Dutch demand], ...
   (PEPYS-E3-P2,8,327.181)

b. Infinitival clauses
   (i) [To groom a horse properly] requires a considerable amount of time, and much skill and exertion;
   (FLEMING-1886,95.517)
   (ii) It is needless [to prove that this idea must be very hurtful to those who entertain it];
   (FROUDE-1830,2,44.294)
c. Participial clauses
   (i) [Polishing and enriching their tongue], is no small business among them;
   (BARCLAY-1743,105.370)
   (ii) It is very trying [governing in a school].
   (THRING-187X,224.239)

In (3-a), there is an alternation between the preposed clausal subject construction and the *it*+subclause construction with an interrogative clause (in this case a *whether*-clause). In (3-b) and (3-c), we find the same alternation with infinitival clauses and participial clauses, respectively.

With respect to the historical corpora, the present study only concerns *that*-clauses, *wh*-clauses (including *whether*-clauses) and infinitival clauses. Participial clauses are excluded from the discussion, as they only begin to develop verbal properties in the course of the Middle English period (Fanego, 2004: 325).

Types of clauses that only occur in the *it*+subclause construction, and not in the preposed clausal subject construction, are not included in the present investigation. Consider the sentences in (4).

(4) *If*-clauses
   a. It would be good [if the remainder of the money due to this Bill could be sent by the next].
   (STRYPE-E3-H,180.8)
   b. *[If the remainder of the money due to this Bill could be sent by the next] would be good.
   [constructed]

In (4-a), we see a regular *it*+subclause construction containing an *if*-clause. However, as represented in (4-b), *if*-clauses do not seem to participate in the preposed clausal subject construction. Furthermore, no such sentences are found in the corpora. *If*-clauses are thus not included.

With respect to Present-day English, the study uses a sample of *whether*-clauses from the British National Corpus (BNC). The reason for

investigating sentences containing such clauses, rather than *that*-clauses, *wh*-clauses or infinitival clauses, is discussed in Chapter 3, where the sample of *whether*-clauses is first presented. In short, it is a result of the lack of phrase structure annotation in the BNC, which makes it hard to find the relevant constructions automatically. Searching for *whether*-clauses is a way to ensure that the sample includes a sufficient number of relevant constructions, while still keeping the sample in a manageable size for manual annotation.

1.4 Structure of the dissertation

The present study is divided into four parts: (i) theoretical framework, material and method, (ii) syntax and argument structure, (iii) weight, complexity and information structure, and (iv) conclusions and suggestions for future research.

Within Part I, Chapter 2 initially gives information on the theoretical stance assumed in the dissertation. The theoretical framework used is Lexical Functional Grammar (LFG), which is a constraint-based generative framework of grammar. Applied to diachronic syntax, LFG, with its dissociation between position and function, is well suited to model shifts in surface realisations of underlying grammatical relations (Vincent, 2001). The different modules of LFG, c(onstituent)-structure, f(unctional)-structure, a(rgument)-structure and i(nformation)-structure, are presented, as well as certain principles and concepts within LFG relevant to the matters discussed here. Particular weight is given to explain choices diverging from mainstream LFG, and to establish the theoretical tools used in later chapters. The two most important theoretical tools are the representations of argument structure and information structure, where I make significant departures from what can be called the 'common ground' within the LFG community.

Chapter 3 contains a presentation of the material and method. After a short introduction to corpus-based research, the corpora used in the dissertation are presented. The corpora include the Penn Corpora of Historical English (PCHE), The York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE) and the British National Corpus. The PCHE and YCOE will collectively go under the name of 'the historical corpora', while the BNC will be taken to represent Present-day English. This chapter also includes a discussion of the corpus annotation and tagging provided, in relation to the relevant constructions. In particular, a description is given of the syntactic annotation of the corpora and

how coding queries can be formulated to code for the different aspects discussed.

Part II concerns the syntax and argument structure of the preposed clausal subject construction and the *it*+subclause construction. In Chapter 4, studies discussing the syntax of clausal subjects and the *it*+subclause construction are presented. This is followed in Chapter 5 by a discussion and analysis of the data from the historical corpora.

Part III of the dissertation is concerned with ways in which weight, complexity and information structure influence the choice of construction. Thus, Chapter 6 concerns studies dealing with these factors in relation to clausal subjects and the *it*+subclause construction. Chapter 7 provides a discussion of the data from both Present-day English and the historical corpora.

In Chapter 8, finally, the results and discussion of the four preceding chapters are summarised and tentative conclusions are drawn. Some suggestions for future research are also presented.

Part I

Theory, material and method

Chapter 2

Theoretical framework

The theoretical framework within which this dissertation is situated is Lexical Functional Grammar (henceforth LFG). LFG is a constraint-based generative theory of syntax, which is formalised in a parallel correspondence architecture, where different levels of grammar (constituent structure, functional structure, phonological structure, information structure etc.) are mapped onto each other. The account of LFG presented in this section is primarily based on Bresnan et al. (2016)[1].

2.1 Background

LFG arose as a theoretical framework for generative linguistics in the early 1980s. It was developed in relation to a discussion (Bresnan, 1977, 1978) about the generality, psychological plausibility and computational suitability of the the hegemonic transformational approach to generative linguistics (e.g. Chomsky 1957, 1965 etc.). Kaplan & Bresnan (1982) show how it is possible to construct a mathematically rigorous account of a range of phenomena, such as raising, control, the active-passive alternation and long-distance dependencies, without the use of transformations.

LFG is a constraint-based linguistic theory. Constraints are set up on the formation of and mapping between different levels of linguistic structure. For the purposes of this study, four levels are particularly important. These are (i) c(onstituent)-structure, (ii) f(unctional)-structure, (iii) a(rgument)-structure, and (iv) i(nformation)-structure. Other levels commonly recognised are p(honological)-structure, m(orphological)-structure and s(emantic)-structure (Dalrymple, 2001).

[1] Frequent references will also be made to Dalrymple (2001).

In the following subsections, a number of aspects of LFG that are relevant for this study are presented. A lot of the theoretical points discussed are uncontroversial and commonly agreed upon. There are, however, a few instances, where theoretical stands are taken, which do not represent the common ground within LFG. These instances are pointed out, as we go along. First, however, let us consider a point which is uncontroversial within LFG, namely the distinction between two levels of syntactic information: c(onstituent)-structure and f(unctional) structure.

2.2 Two levels of syntactic information

LFG assumes two different ways of representing syntactic information. Phrasal constituent structure is separated from abstract functional syntactic concepts such as predicate-argument structure and concepts such as subject and object. The phrasal constituent structure is called c-structure and is typically represented in the form of a syntactic tree. The abstract functional syntactic organisation is known as f-structure and is typically represented in the form of an attribute-value matrix (an AVM).

In this section, the most relevant aspects of first c-structure and then f-structure are presented.

2.2.1 C-structure

C-structure represents phrasal constituent structure and the way in which phrases and words can be substituted for each other, moved around and deleted. An important point concerning c-structure is that the rules governing this level of syntax, the phrase-structure rules, are assumed to be specific for each language. Categories and configurations are only present if there is evidence from that particular language that a category is warranted. Consider the English sentence in (1). The c-structure of (1) could be represented as in 2.1.

(1) John often eats fish.

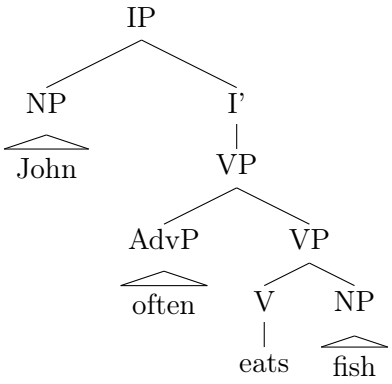

Figure 2.1: C-structure for the sentence 'John often eats fish'

The figure Figure shows the phrasal constituent structure of the sentence in (1). As can be seen, we see the familiar tree with an IP and a VP. This structure holds for English. In other languages, the tree might look different.

For Swedish, a translation of the sentence in (1) would be the sentence in (2). The f-structure for (2) is given in 2.2.

(2) Johan äter ofta fisk.
John eats often fish
'John often eats fish'.

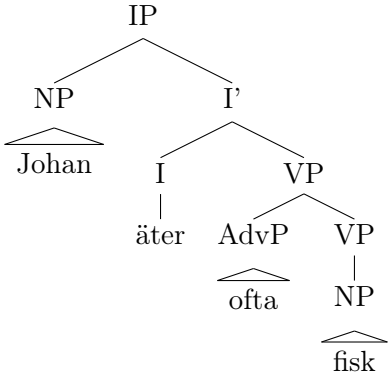

Figure 2.2: C-structure for the sentence 'Johan äter ofta fisk'

As can be seen by comparing the c-structures in 2.1 and 2.2, there is a difference with respect to the position of the finite lexical verb. Assuming that the ADJ(unct) *often/ofta* has the same position in the two languages,

we can see that the English finite lexical verb follows the adjunct while, in Swedish, it precedes it. This, in conjunction with other differences, can be taken as evidence to suggest that the position of the verb differs between the two languages. One point of interest, which should be commented on, is that, in the tree in 2.2, we have a VP constituent without a verb (cf. Nordlinger & Bresnan, 2011: 121). This is a generalisation, based, for instance, on the fact that the lexical verb and the object constituent do form a constituent when another element fills the head of the IP phrase.

(3) Äter fisk gör Johan ofta.
eats fish does John often
'John often eats fish'.

In (3), the lexical verb *eats* and the object *fish* form a constituent when the element *gör* ('does') constitutes the head of I.

The c-structure is governed by a number of principles. I will here present two such principles: (i) Economy of expression and (ii) Lexical integrity. These are given in (4) and (5)

(4) Economy of Expression:
All syntactic phrase structure nodes are optional and are not used unless required by independent principles (completeness, coherence, semantic expressivity). (Bresnan et al., 2016: 90)

(5) Lexical Integrity:
Morphologically complete words are leaves of the c-structure tree and each leaf corresponds to one and only one c-structure node. (Bresnan et al., 2016: 92)

The economy of expression principle says that all nodes in a tree are optional unless required by some independent principle. Recall for instance that the V node in the tree in 2.2 is absent as there is no principle that requires it to be there. The second principle, the Lexical integrity principle, means that there is a strict separation between syntax and morphology. As opposed to transformational grammar, the c-structure rules in LFG only concern morphologically complete words (cf. Dalrymple, 2001: 84). A consequence of the lexical integrity principle, in conjunction with the economy of expression principle, is that there are no 'empty' nodes or nodes associated with phonologically unrealised constituents. This is highly relevant for the constructions analysed in the present study, as there is no c-structure realisation of null subjects (such as pro/PRO, within Minimalism).

As can be seen, there are several phenomena, such as case or grammatical functions (e.g. subject and object), which are not represented in the c-structure of LFG. In Minimalism, these (e.g. case) are often represented in the form of a syntactic projection in the constitutent tree. In LFG, such abstract syntactic information is represented in f-structure, which is the topic of the next section.

2.2.2 F-structure

The f-structure contains information about grammatical functions such as subject and object, as well as other morphosyntactic information such as tense, aspect, mood, case, number, etc. The relationship between f-structure and c-structure is governed by a function known as the ϕ function, which maps c-structure nodes onto f-structures. The f-structure that corresponds to the c-structure representation in Figure 2.1 is represented in Figure 2.3. This is a simplified f-structure, for expository purposes.

$$\begin{bmatrix} \text{PRED} & \text{`eat } \langle \text{SUBJ, OBJ} \rangle\text{'} \\ \text{TENSE} & \text{PRESENT} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{`John'} \end{bmatrix} \\ \text{OBJ} & \begin{bmatrix} \text{PRED} & \text{`fish'} \end{bmatrix} \\ \text{ADJ} & \begin{bmatrix} \text{PRED} & \text{`often'} \end{bmatrix} \end{bmatrix}$$

Figure 2.3: F-structure for the sentence 'John often eats fish'

The f-structure in Figure 2.3 represents information about the grammatical functions, SUBJ(ect), OBJ(ect) and ADJ(unct), as well as information about TENSE. It is formalised in an attribute-value matrix, where attributes, such as TENSE, are mapped to values, such as PRESENT. While the c-structure representations of the English and Swedish sentences in (1) and (2) differ from each other, the f-structural information is the same, i.e the f-structural information associated with the sentence *John often eats fish* is the same information as for the Swedish sentence *Johan äter ofta fisk*, despite the fact that the lexical items have different forms. The so-called PRED feature seen in Figure 2.3 is a peculiarity of f-structure. It gives syntactically relevant semantic information about the f-structure constituent, such as the subcategorisation of a predicate (e.g. 'eat ⟨SUBJ, OBJ⟩') or the lexical identity of a constituent (e.g. 'John').

In the same way as for the c-structure, there are principles governing the formation of the f-structure. This level of syntax is governed by two well-formedness conditions: (i) completeness and (ii) coherence. Completeness says that every function designated by a PRED must be present in the f-structure and coherence says that every argument (i.e. not adjunct) function must be designated by a PRED (Bresnan et al., 2016: 17). A consequence of these two conditions is for example that a SUBJ, in order to be licensed, must be designated by a PRED. As will be seen, this is relevant for the analysis of expletives and their relation to different kinds of predicates.

2.2.3 Formulation of syntactic constraints

As laid out in the two previous sections, syntactic information is divided up into two different representations, c-structure and f-structure. In the formulation of descriptions and constraints on c-structure, f-structure, and the relation between them, LFG makes use of lexical entries, phrase structure rules, and so-called f-descriptions. Lexical entries represent the information associated with a lexical item. The lexical items *fish* and *eats* could be represented as in (6).

(6) a. Lexical entry for the item *fish*:

fish	N	(↑ PRED) = 'fish'
		(↑ NUM) = NONCOUNT
		...

b. Lexical entry for the item *eats*:

eats	V	(↑ PRED) = 'eat ⟨SUBJ, OBJ⟩'
		(↑ TENSE) = PRESENT
		...

The f-structural information that is entered for the lexical item *fish* here is that it is a noun, that it has the PRED value 'fish' and that it is noncountable. For the verb *eats*, the f-structural information is that it is a verb, that it has the PRED value 'eat ⟨SUBJ, OBJ⟩', and that it is in the present tense. Shortly, we will see more infromation that could be entered for the lexical item *eats*. All the lexical items in a language are associated with a lexical entry such as the ones in (6).

The descriptions and constraints on combining lexical items is formulated in phrase structure rules. One possible oversimplified rule for verb phrases with monotransitive verbs such as *eat* in English is the one in (7) (Dalrymple, 2001: 94).

(7) Simplified verb phrase rule:

VP → V NP

The rule in (7) says that the admissable daughters of a VP are a V in conjunction with an NP.

In the same way as phrase structure rules are used to formulate constraints on c-structure, f(unctional)-desciptions can be used to formulate constraints on f-structure. The f-descriptions in (8) state that the SUBJ (subject) of the f-structure of the mother node is required[2] to have the 3rd person singular morphology. It could be entered in the lexical entry of *eats*, stating that the SUBJ of *eats* is in the 3rd person singular.

(8) F-descriptions:

$(\uparrow \text{SUBJ NUM}) =_c \text{SG}$
$(\uparrow \text{SUBJ PERS}) =_c \text{3RD}$

The symbol ↑ in (8) is a way of locating the f-structure associated with the SUBJ in a c-structure tree. The symbol ↑ refers to the f-structure of the mother node and the ↓ refers to the f-structure of the self node (Dalrymple, 2001: 118). An example of how f-structures are located using these symbols is given in (9).

(9) Annotated phrase structure rule:

IP → NP I'
$(\uparrow \text{SUBJ}) = \downarrow$ $\uparrow = \downarrow$

The annotated phrase structure rule in (9) states that the NP daughter of the IP provides the SUBJ of the IP. The I' daugther of the IP provides whatever f-structure information that its daughter(s) provide(s). Annotated phrase-structure rules such as the one in (9) are used to formulate constraints on the c-structure-to-f-structure correspondences. Information from other linguistic levels, such as a-structure or i-structure, can also be incorporated into annotated phrase structure rules in order to formulate various constraints.

In the chapter so far, information has been given on relevant aspects of c-structure, f-structure and correspondences between these two levels. In the next section, we need to elaborate on an important aspect of f-structure, namely the notion of the grammatical function. In Minimalism, concepts like subject- or objecthood follow from the position of phrases in the

[2]The subscripted c in (8) indicates that the equation is a constraining equation which requires a certain feature to be present.

constituent structure. In LFG, these notions are atomic and independent. Grammatical functions are discussed in more detail in the next section.

2.3 The grammatical functions

As said above, the grammatical functions, such as subject and object, have a particular status in LFG in the sense that they are atomic concepts that do not follow from the constituent structure. Based on Asudeh (2012), I will assume the grammatical functions in (10).

(10) Argument functions: SUBJ, OBJ, OBJ_θ, OBL_θ.
Adjunct function: ADJ.
Unbounded dependency function: UDF[3].

Argument functions (AFs) are distinguished from adjuncts. We also have the unbounded dependency function UDF. For the moment, let us focus on the argument functions. Apart from the transparent labels SUBJ and OBJ, we have two additonal argument functions, OBJ_θ and OBL_θ. The function OBJ_θ represents semantically restricted objects, such as the secondary object in a ditransitive construction, which is restricted to the role of recipient. The function OBL_θ represents semantically restricted oblique functions, i.e. lexically selected non-internal arguments that are restricted to certain semantic roles. One example of an OBL_θ is the locative argument of a verb such as *live* in the sentence *I live in Gothenburg*, which would be classified as an oblique function restricted to the thematic role of location, i.e. OBL_{loc}.

There is an on-going discussion within the LFG community on the nature of the grammatical functions. In the next two subsections, some of the stands I take with respect to the grammatical functions that are relevant for the present investigation are explained and discussed. The first subsection concerns the SUBJ function, and the second subsection clausal complements.

2.3.1 The SUBJ function

Let us first consider the properties generally associated with subjecthood, and then proceed to the representation of the SUBJ function within

[3]The most common position within LFG is to make use of the so-called grammaticalised discourse functions (GDFs) TOP(ic) and FOC(us) (e.g. Falk, 2001: 60). However, given the fact that I assume a separate representation of information structure, I use the function UDF, which replaces both TOPIC and FOCUS in the f-structure (cf. Asudeh, 2012: 72).

LFG. In modern descriptive syntax (Givon, 2001), there are a number of properties associated with subjecthood, not all of which are present in all languages. Givón (2001) gives a list of such properties. She divides the subject properties into (i) overt coding properties and (ii) behaviour-and-control properties. The overt coding properties of subjects include the structural position of the subject vis-a-vis other GFs and the verb, verb agreement and nominal morphology. The behaviour-and-control properties of subjects include raising, passivisation, reflexivisation and anaphoric co-reference in chained clauses.

Although the above properties of subjecthood are generally agreed upon, there are considerable differences in the way in which subjects are represented within formal theories of syntax. As described in the introduction, LFG here differs from for instance the Minimalist Program. Subjecthood within Minimalism is a structural property that can be derived from the structural position of the constituent. Within LFG, the subject function (as well as the other grammatical functions) constitutes a theoretical primitive, which cannot be derived from other theoretical constructs. There is thus no inevitable connection between subjecthood and structural position. In languages where there is a structural subject position, like English, such a constraint can be formulated, but it is not a cross-linguistically necessary property of subjects to occur in a particular structural position.

Apart from having the properties outlined by Givón (2001), the SUBJ function furthermore has always had a special status within LFG with respect to argument structure. Unlike the other grammatical functions, a constraint is commonly assumed that all predicates require a subject. This constraint is called the Subject Condition. Bresnan et al. (2016: 334) formulate the condition as follows: 'every predicator must have a subject'. Dalrymple (2001: 19) refers back to Bresnan & Kanerva (1989) and gives the slightly more specific 'every verbal predicate must have a SUBJ'. In the theory of argument structure assumed here (Kibort, 2007), the Subject Condition is rejected, in opposition to mainstream LFG. The assumptions made here with respect to argument structure and the rejection of the Subject Condition are further discussed under the heading Lexical Mapping Theory in Section 2.4.

2.3.2 Clausal complements

In LFG, it is often assumed that all clausal and verbal complements express the grammatical functions COMP and XCOMP (e.g. Bresnan et al., 2016: 99), respectively, which are specifically designated for clausal

and verbal complements (for a discussion on nominal COMP, see Lødrup, 2012). There are different ways of reconciling the idea that all clausal and verbal complements are COMPs or XCOMPs with the set of argument functions listed in (10), SUBJ, OBJ, OBJ_θ, OBL_θ. If you do not want to posit additonal argument functions, a popular solution is the one presented in Zaenen & Engdahl (1994), where COMP and XCOMP are equated with the function OBL_{prop}, i.e. an oblique function restricted to the semantic role of proposition. This is also the solution adopted for the present investigation, where I will use the labels COMP and XCOMP for the closed and open[4] versions of the function OBL_{prop} (cf. Falk, 2005).

The idea of a specifically designed function for clausal complements is based on the generalisation that clausal complements behave differently syntactically from nominal complements. Based on Huddleston & Pullum (2002: 1017-1021), I will here mention three ways in which clausal complements differ from nominal complements: (i) linear position, (ii) lexical choice, and (iii) being the complement of a preposition. As will be seen, these observations are not equally relevant for the choice to adopt the grammatical function COMP. The first thing to be discussed is linear position. Example (11) shows that clausal arguments follow a manner adverb in a situation where a nominal complement typically would precede it.

(11) a. ?He opened slowly the door.
b. He denied categorically that he had spoken to her.
(Huddleston & Pullum, 2002: 1018)

In (11), the first sentence where the nominal complement follows the manner adverb *slowly* is questionable at best. The second sentence, on the other hand, where there is a clausal complement following the manner adverb, is perfectly fine. Huddleston & Pullum (2002) presents this as a difference between clausal and nominal arguments. It is not clear that such a difference in word order should be relevant for the adoption of the grammatical function COMP. Linear order is a c-structure property, which is not sufficient to determine the status of grammatical functions.

The second piece of evidence is that there are verbs, such as *hope*, that only take clausal arguments and not nominal arguments. In these cases, the clausal argument alternates with an oblique prepositional phrase.

(12) I hope [that it will rain] / *it / *(in) your words.

[4]The grammatical function XCOMP is a so-called open complement function, used in connection with functionally controlled infinitives and participles.

As can be seen in (12), the verb *hope* takes a *that*-clause as a complement, but neither the pronoun *it* nor the noun phrase *your words* can be used as complements to this verb, unless the noun phrase functions as the complement to a preposition. Considering the fact that subcategorisation is represented at f-structure, the alternation between clausal complements and prepositional phrases is a strong argument that the clausal complement should be analysed as a COMP (=OBL_{prop}) rather than OBJ.

Thirdly, *that*-clauses in English do not occur as complements to prepositions, while nominal elements do. Consider (13). *Wh*-clauses, however, do occur as complements to prepositions.

(13) a. He rejoiced at her decisive victory
b. *He rejoiced at that she had won so decisively.
(Huddleston & Pullum, 2002: 1019)

Here, we see that the NP *her decisive victory* functions as the complement of the preposition *at*, while the *that*-clause *that she had won so decisively* is ungrammatical in the same environment. This argument follows up on the earlier argument about the alternation between clausal complements and prepositional phrases.

In summary, there are arguments supporting the distinction between the two complement functions OBJ and COMP (see Alsina et al. (2005) for a different analysis). Especially important is the alternation between subclauses and prepositional phrases in environments where nominal constituents are not possible. Since prepositional phrases typically express oblique functions, it makes sense that the clausal complements here alternate with prepositional phrases.

Further support comes from the interaction between the grammatical functions and the thematic roles described within the mapping theory which is outlined in the next section. Compare the possible complements of the verb *believe* in (14) with the previously given complementation pattern of the verb *hope* in (12).

(14) I believe [that the earth is round] / so / it / (in) the Prime Minister.

For the verb *hope* in (12), all complements consisting solely of a noun phrase are excluded, while for the verb *believe*, noun phrases, clauses and prepositional phrases are all possible as complements. This difference between the two verbs can be tied to the interaction between thematic roles and grammatical functions. In the mapping theory, a distinction is made between the possible grammatical functions mapped to by the

thematic role *theme* and the possible grammatical roles mapped to by the thematic role *proposition*. For the verb *hope*, it seems that this verb consistently takes the thematic role *proposition* as an argument, which in the mapping theory adopted here is mapped to COMP. The verb *believe*, on the other hand, have two different thematic role selections, depending on its interpretation. Consider the sentences in (15), along with their respective interpretations.

(15) a. I believe the Prime Minister / it = 'I trust what the Prime minister has said'.
b. I believe [that the earth is round] / so[5] = 'I hold the proposition for true that the earth is roundl'.

In (15), we see two different interpretations of the verb *believe*, $believe_1$ and $believe_2$. In (15-a), the verb *believe*, $believe_1$, has the approximate meaning 'to put your trust in'. It then takes a takes a locutionary act, an actual utterance as a complement. This complement has the thematic role *theme* and the grammatical function OBJ. In (15-b), on the other hand, the verb *believe*, $believe_2$, has the approximate meaning 'to hold a proposition for true'. In this case, the complement has the thematic role *proposition* and the grammatical function COMP. The differences in complementation pattern between the verbs *hope* and *believe*, according to the account developed here, follow from the mapping between thematic roles and grammatical functions for these verbs. The verb *hope* and $believe_2$ ('hold a proposition for true') take a *proposition* as a complement, which is mapped to the grammatical function COMP. The other interpretation of the verb *believe*, $believe_1$, has a different thematic role selection, where the complement has the thematic role *theme*, which is mapped to the grammatical function OBJ.

In this section, the function COMP and its relation to the other grammatical functions have been discussed. It is concluded that it is convenient to analyse certain clausal complements as COMP (=OBL_{prop}) to account for differences in complementation patterns between verbs in English. After thus having discussed the relevant grammatical functions, the next section concerns the way in which these grammatical functions are connected to the thematic roles they are associated with through the system of Lexical Mapping Theory.

[5]There is a difference between the pro elements *so* and *it* in relation to the verbs *hope* and *believe*, which corresponds to the difference between nominal and clausal complements. The verb *hope* takes *so*, but not *it*, while the verb *believe* takes either pro element. One possible analysis of the pro element *so* would be that it is a pronominal OBL, while *it* is a pronominal OBJ (or SUBJ).

2.4 Lexical Mapping Theory

As mentioned, in contrast to for instance the Minimalist Program, grammatical functions such as subjects and objects are theoretical primitives in LFG. Being a subject or object thus does not follow from the structural position of thematic participants such as agent and patient. Instead, the relation between thematic roles and grammatical argument functions (AFs) is governed by mapping rules in a theory known as Lexical Mapping Theory (LMT). The thematic roles and grammatical functions are mapped to each other through an intermediate layer of representation known as a(rgument)-structure. The relation between these three levels, s-structure, a-structure and f-structure can be schematically represented as in Figure 2.4.

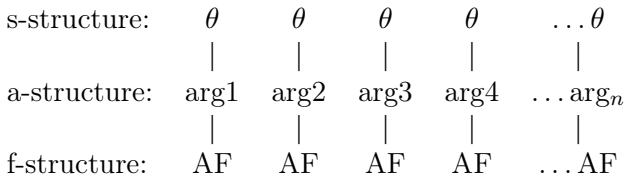

Figure 2.4: Mappings between s-structure, a-structure and f-structure

Thematic roles represent the participants in the event denoted by the verb. Examples of semantic participants are agent, beneficiary, experiencer, instrument etc. These participants are in LFG mapped to the argument positions, arg_n, of a-structure. The respective argument slots constitute the link between thematic roles and grammatical argument functions. Argument functions represent the syntactic arguments selected by a verb. The argument functions used here are SUBJ, OBJ, OBL_θ and OBJ_θ.

An important insight, which forms the basis of Lexical Mapping Theory, is that the grammatical functions form natural groups based on their behavior. For example, the SUBJ and OBJ functions can be grouped together based on the fact that they are semantically unrestricted. They are unrestricted in the sense that they can be filled by an argument with any thematic role or by an expletive. Within LMT, two features are commonly used to categorise the grammatical functions into groups. These two features are (i) restrictedness (whether or not the GF is semantically restricted) and (ii) objecthood (whether or not the GF is an internal argument of the predicate). The two features, [±r] and [±o], form the basis for the the classification in Figure 2.5 (Bresnan, 2001).

	−r	+r
−o	SUBJ	OBL_θ
+o	OBJ	OBJ_θ

Figure 2.5: Cross classification of argument functions.

The classification in Table 2.5 is made use of in the mapping rules. It also makes it possible to rank the grammatical functions in a markedness hierarchy. The SUBJ function is unrestricted semantically and is also non-objective, which means that it is the least marked GF. The OBJ_θ function, on the other hand, is both restricted and objective and thus the most marked GF. The markedness hierarchy is represented in (16).

(16) Markedness hierarchy
SUBJ ≻ OBJ, OBL_θ ≻ OBJ_θ

In one of the most widely adopted versions of LMT, the one presented in for instance Bresnan (2001), thematic roles are assigned an intrinsic feature (either ±r or ±o), which governs what group of GFs this thematic role can be mapped to. In the present study, instead of the approach to the Lexical Mapping Theory presented in Bresnan (2001), I adopt the revised Lexical Mapping Theory of Kibort (2007, 2008, 2013, 2014). Kibort (2007) assigns intrinsic features to argument slots rather than to thematic roles. For a discussion of the general benefits of this approach, see Kibort (2007). This means that we have an independent level of a-structure where a number of argument positions are represented together with their intrinsic features. The a-structure is represented in Figure 2.6.

arg1,	arg2,	arg3,	arg4,	$\ldots \text{arg}_n$
[–o]/[–r]	[–r]	[+o]	[–o]	[–o]

Figure 2.6: Intrinsic classification of a-structure positions

As described in Kibort (2014), the semantic participants of an event are restricted as to which argument position(s) they can be mapped to. An agentive participant is for instance typically restricted to the arg1[–o] slot. Using the formalisation in Kibort (2014), a participant which can only be mapped to the arg1 slot has the semantic marker **1** in the entry for that particular predicate. Certain semantic participants can be mapped to more than one argument position. There could for instance be a participant which can be mapped to either arg2 or arg3, and which therefore would have the semantic marker **23**. The choice of argument position, when a participant can be mapped to more than one argument position, is not specified in Kibort (2014), but is assumed to be the result of the lexical semantics of the predicator in a particular context. For the purpose of exposition, the semantic markers will not be shown, and two different argument structures will be presented for predicates where the semantic participants can be mapped to more than one argument slot.

The argument positions represented in Figure 2.6 are in turn mapped to the grammatical functions of f-structure. This mapping between argument slots and AFs is governed by the principle in (17).

(17) Mapping principle:
The ordered arguments are mapped on to the highest (i.e. least marked) compatible function on the markedness hierarchy

Let me give an example of how the mapping works. Consider the active-passive alternation in (18).

(18) a. John beat Tom.
[constructed]
b. Tom was beaten by John.
[constructed]

The mapping between thematic roles and grammatical functions can be represented as in (19).

(19) Argument-to-function mapping for the verb *beat*:

	agent	*patient*
	|	|
beat	⟨ arg1,	arg2 ⟩
	[–o]	[–r]
	|	|
	⟨SUBJ,	OBJ⟩

The verb *beat* has two semantic participants, an *agent* and a *patient*. The *agent* participant has the semantic marker **1** and the patient has the semantic marker **2**. The *agent* role is thus mapped to arg1[–o] and the *patient* role to arg2[–r]. The least marked compatible AF for the arg1[–o] position to be mapped to is SUBJ. When the SUBJ function is taken, the least marked compatible function for the arg2[–r] position is the OBJ function. Thus, we get the mapping in (19).

The default argument-to-function mapping shown above can be interfered with by so-called morphosyntactic operations (Kibort, 2014). Morphosyntactic operations alter the argument-to-function mapping without affecting the lexical or semantic tiers of representation, i.e. they are meaning preserving. Kibort (2014) assumes three morphosyntactic operations, shown in (20).

(20) Morphosyntactic operations:

a. adding the [+r] specification to a [–o] argument;
b. adding the [+o] specification to a [–r] argument; and
c. adding the [+r] specification to a [+o] argument.

In the case of the passive sentence in (18-b), the argument-to-funtion mapping is the result of the morphosyntactic operation in (20-a). Consider the mapping in (21).

(21) Argument-to-function mapping for the passive participle *beaten*:

	agent	*patient*
	|	|
beaten	⟨ arg1,	arg2 ⟩
	[–o]	[–r]
	[+r]	
	|	|
	⟨OBL$_\theta$,	SUBJ⟩

In (21), the arg1[–o] position is assigned an extra [+r] feature. The result of this operation is that the arg1[–o] argument is forced to map

to a semantically restricted grammatical function, namely the OBL_{agent} function, which is the least marked and only function compatible with the arg1[–o, +r] slot. For the arg2[–r] position, the SUBJ function is now up for grabs, being the least marked compatible AF.

The argument structure analysis of the passive and the morphosyntactic operations play an important role in the analysis of preposed clausal subjects and the *it*+subclause construction in part II, Chapter 5.

With the above presentation of the Lexical Mapping Theory, the most important aspects of c-structure, f-structure and a-structure have been described and discussed. In the following and final section of the theory chapter, we turn to the representation of information structure.

2.5 Representation of information structure

Information structure (IS) concerns the ways in which semantic content, which in LFG is represented at s(emantic)-structure, is packaged in various linguistic form depending on the speaker's assumptions about the current information state[6] in a discourse (cf. Chafe, 1976). Consider the utterance in (22). The small caps in the example indicate a contrastive main accent. The rest of the utterance is destressed.

(22) He saw MARY.
[constructed]

The form of the utterance in (22) gives rise to a number of assumptions on the part of the speaker/hearer with respect to information structure. These assumptions concern the activation and accessibility of the discourse referents, i.e. to what extent a discourse referent is assumed to be active in the mind of the addressee, and givenness, i.e. whether some proposition is assumed to be part of the addressee's information state before an utterance is made or whether the proposition enters the information state of the addressee as a result of an utterance. In the utterance in (22), the status of the subject *he* as a pronoun, rather than a full noun phrase, signals that the referent associated with *he* is assumed to be identifiable and active in the mind of the addressee (cf. Gundel et al., 1993). Furthermore, the status of *Mary* as a proper name, signals that the referent is assumed to be at least identifiable. With respect to givenness, assuming that the the subject and the finite verb are destressed, while the main accent of the

[6]The term *information state* commonly refers to the mindset of the participants in a discourse including established discourse referents as well as the propositions that the participants share (Krifka & Musan, 2012: 1).

intonation phrase is on the object, the proposition that *John saw someone* is assumed to be part of the information state of the addressee before the utterance (i.e. given), while the proposition that *John saw Mary* is added to the information state of the addressee after the utterance is made (i.e. new). Lastly, assuming that there is a contrastive pronunciation of *Mary*, i.e. if the pronunciation of *Mary* shows greater phonetic prominence (pitch, duration and intensity) than what it would have had as discourse-new (Katz & Selkirk, 2011), there is assumed to be a contrastive relation between the element *Mary* and some other referent. The information structural properties of the utterance in (22) constrains what questions this utterance could constitute the answer to. The question corresponding to the information structural properties of the utterance in (22) is given in (23).

(23) Did John see Mary or someone else?
[constructed]

With respect to the formal representation of information structure within LFG, following work such as King (1997); Choi (1999); Dalrymple & Nikolaeva (2011), I assume an independent level of representation called i(nformation)-structure, where the meanings from the s(emantic) structure are ordered according to their information structural properties. The features and properties I assume for the present study are given in 2.7, which is a slightly modified version of those assumed in Dalrymple & Nikolaeva (2011).

$$\begin{bmatrix} \text{STATUS} & \{\text{IDENTIFIABLE, UNIDENTIFIABLE}\} \\ \text{ACTV} & \{\text{ACTIVE, INACTIVE, ACCESSIBLE, ANCHORED}\} \\ \text{GIVENNESS} & \{\text{GIVEN, NEW}\} \\ \text{CONTRAST} & \{\text{CONTRASTIVE, NONCONTRASTIVE}\} \end{bmatrix}$$

Figure 2.7: Features and values at i-structure

Figure 2.7 shows four attributes: (i) STATUS (identifiability status), (ii) ACTV (activation of dicourse referent), GIVENNESS (givenness relation) and CONTRAST (presence of a contrastive relation). The attribute STATUS has the values identifiable and unidentifiable (i.e. whether or not a dicourse referent can be assumed to be identified by the addressee). The attribute ACTV has the values ACTIVE, INACTIVE, ACCESSIBLE, ANCHORED (depending on the assumed activation in the mind of the addressee of a

discourse referent. The attribute GIVENNESS has the two values GIVEN and NEW, concerning the assumed presence or absence of a proposition in the information state of the addresee. Lastly, the attribute CONTRAST has the two values CONTRASTIVE and NONCONTRASTIVE, concerning the presence of an assumed contrastive relation between an element in an utterance and other contextually available elements.

In relation to the utterance given in (22), we have seen that the form of this utterance gives rise to a number of assumptions with respect to information structure. A formal representation of these information structural assumptions is given in Figure 2.8, based on the features in Figure 2.7. The elements in bold face represent meanings from the s-structure.

$$\begin{bmatrix} \text{IDENTIFIABLE} & \{\textbf{john}, \textbf{mary}\} \\ \text{ACTIVE} & \{\textbf{john}\} \\ \text{GIVEN} & \{\textbf{john saw x}\} \\ \text{NEW} & \{\textbf{john saw mary}\} \\ \text{CONTRASTIVE} & \{\textbf{mary}\} \end{bmatrix}$$

Figure 2.8: I-structure for the sentence *He saw* MARY.

The figure shows that the referent of *John* is IDENTIFIABLE and ACTIVE, while the referent of *Mary* is IDENTIFIABLE. The proposition that *John saw someone* is GIVEN, while the proposition that *John saw Mary* is NEW. Furthermore, the dicourse referent of *Mary* is CONTRASTIVE. It is important to emphasise that the information structure in 2.8 follows exclusively from the syntactic and phonological properties of the sentence uttered in (22). It does not follow from the properties of the context.

In Chapter 7, an analysis will be presented of the preposed clausal subject construction in relation to the *it*+subclause construction based on an operationalisation of the information structural concepts of this section.

2.6 Summary

In this chapter, the theoretical assumptions that provide the point of departure for the present investigation have been presented and discussed. It has been shown how the architecture of LFG is based on a parallell correspondence between different levels of linguistic information. The

c-structure provides information about the phrasal constituent structure. The f-structure provides information about abstract syntactic relations such as agreement, grammatical functions (e.g. subject and object), syntactic features such as case, tense, gender and number. A-structure represents sub-categorisation information and could be taken to work at the interface between f-structure and s-structure. The mapping between s-structure, a-structure and f-structure is governed by the principles laid out in the Lexical Mapping Theory. Lastly, the i-structure has been presented where the meanings from the s-structure are ordered with respect to their information structural features and roles.

The next chapter deals with the method and material used in the present study.

Chapter 3

Material and method

The present investigation is corpus-based. It is so in the sense given by Tognini-Bonelli (2001) that electronically stored text (assembled in order to provide a representative sample of a language variety) is employed to describe and explain linguistic patterns of variation and use. In the present chapter, the corpora chosen for this dissertation are presented as well as the annotation and tagging used in the corpora. Some space is also given to explain the search program CorpusSearch and the coding queries made in the attempt to investigate the phenomena of the dissertation.

3.1 Material

The corpora that provide the material for the present investigation are listed in the following:

- The York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE)
- Penn-Helsinki Parsed Corpus of Middle English (PPCME2)
- Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME)
- Penn Parsed Corpus of Modern British English (PPCMBE)
- The British National Corpus (BNC)

The corpora can be divided into three groups. First, we have the York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE), which represents the earliest stage of the English language, Old English (-1100). Then we have the three corpora collectively known as the Penn Corpora

of Historical English (PCHE), which together cover the period from about 1100 until the beginning of the first World War. Lastly, we have the British National Corpus (BNC), which contains British English material collected in the 1980s and 1990s. The corpora, except for the BNC, are devised in such a way that they are comparable to each other. The Old English prose corpus and the Penn Corpora of Historical English contain texts from approximately the same text genres[1], such as law, science, homilies, religious treatises, history, biographies, rules and bible passages, and contain approximately the same number of words, between approximately 1 million and 1.7 million words. The BNC is a considerably larger corpus, about 100 million words, which to a certain extent contains other types of material. The sizes of the corpora are given in Table 3.1.

Table 3.1: Size of the corpora.

Corpora	Size (in number of words)
Old English prose (YCOE)	1.5 million
Middle English (PPCME2)	1.2 million
Early Modern English (PPCEME)	1.7 million
Late Modern English (PPCMBE)	0.95 million
Present-Day English (BNC)	100 million
Total	105.35 million

In the following, the corpora are described in more detail.

3.1.1 The Old English prose corpus

The York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE) contains prose texts from the earliest period of the English language between 600 and 1100. The corpus contains 1.5 million words of syntactically annotated Old English texts, which represents approximately half of all preserved material of Old English (http://www.doe.utoronto.ca/). It consists of 100 texts of various lengths, from 100 words to 100,000 words. The distribution of the texts within different genres is shown in Table 3.2[2].

As can be seen in Table 3.2, there are three genres that are particularly dominant: (i) homilies, (ii) biographies and (iii) histories. Together, these three genres represent approximately 64 % of the corpus.

[1]The text genres of diaries and personal letters are not included in the Old and Middle English corpora.

[2]The percentages are rounded to one decimal place.

Table 3.2: Texts in the Old English prose corpus

genre	No of texts	No of words	%
Homilies	8	345,251	23.8
Biography, lives	19	343,990	23.7
History	6	236,165	16.3
Bible	4	136,948	9.4
Religious treatise	18	129,993	9.0
Handbooks, medicine	4	68,315	4.7
Philosophy	2	50,623	3.5
Rules	2	38,490	2.7
Laws	10	20,807	1.4
Apocrypha	5	19,867	1.4
Science	2	15,738	1.1
Charters and wills	6	11,906	0.8
Ecclesiastical law	4	11,309	0.8
Travelogue	1	7,271	0.5
Fiction	1	6,545	0.5
Preface	6	4,302	0.3
Geography	1	1,891	0.1
Epilogue	1	965	0.1
total	100	1,450,376	100

A large proportion of the texts of the corpus have a religious theme, such as homilies, biographies of saints, translations of bible passages, and religious treatises. This is not at all surprising, given the spread of the Roman alphabet as a result of the spread of the Roman church (Lapidge, 1986: 5).

A consequence of the fact that writing is so intimately connected to the spread of Christianity concerns the use of Latin. Latin was the language used for the written production of the scribes working in the monasteries. As a result of the importance of Latin as the main language of written communication in Europe, it is necessary to ask to what degree the Old English texts we have are influenced by Latin. It turns out that about 50 % of the texts in the Old English prose corpus constitute translations from Latin. As will be seen in Chapter 5, certain syntactic constructions in the Old English texts, in particular the raising construction, are only found in translations from Latin.

The question of the relationship between Latin and Old English comes particularly to the fore in certain texts. The Old English translation of Bede's *Ecclesiastical History* is one text that warrants closer attention. The original text was composed in Latin in the 8th century by the monk Bede, later known as the Venerable Bede (Beda Venerabilis). It was translated into Old English during the reign of King Alfred (871-899), probably by a scribe of Anglian origin (Treharne, 2010: 2).

There are a number of things that make the *Ecclesiastical History* worth extra attention. Firstly, it is one of the longest text in the YCOE, containing approximately 80,000 words. Secondly, it has some grammatical features that are interesting for the current study. One such grammatical feature is the proportion of verb-initial clauses. It seems as if the *Ecclasistical History* has an unusually large proportion of verb-initial clauses in comparison to other Old English texts (Calle-Martína & Miranda-García, 2010). Another interesting thing about the *Ecclesiastical History* is that it, in comparison to other texts, contains a considerable number of propositional subclauses with and without a subject *it*.

The syntax of the *Ecclesiastical History* has been described as 'convoluted' (Treharne, 2010: 2), and it not clear whether the above-mentioned features of the text derive from the fact that the translator is of an Anglian origin, that it is a translation from Latin, or something else. However, the fact that it is a translation from Latin cannot be ignored in the analysis of the syntax of this particular text.

The large proportion of texts being translations from Latin and being associated with a religious theme sets the Old English corpus apart from the other corpora used for the present investigation.

3.1.2 The Middle English corpus

The Penn-Helsinki Parsed Corpus of Middle English, in its second edition, contains around 1.2 million words in 56 different texts, covering the period from around 1100 to about 1500. The genre distribution of the texts is given in Table 3.3.

Just as for the Old English period, it is the religious texts that dominate the scene in the Middle English period; religious treatises, sermons, homilies and bible texts represent approximately 62 % of the corpus. History texts also constitute an important genre. One thing that should be noted is that there is no letter correspondence included in the corpus,

Table 3.3: Texts in the Middle English corpus

genre	No of texts	No of words	%
Religious treatise	17	382,429	33.1
History	5	192,342	16.6
Sermon	7	136,869	11.8
Homilies	7	131,327	11.4
Bible	3	65,617	5.7
Romance	2	65,434	5.7
Travelogue	1	49,690	4.3
Rule	2	35,234	3.0
Handbooks	5	33,328	2.9
Biography, lives	4	27,425	2.4
Philosophy	2	27,420	2.4
Fiction	1	17,005	1.5
total	56	1,155,965	100

even though there is letter correspondence available from late Middle English[3].

The material is not evenly distributed through the time period, as the bulk of the material derives from late Middle English. About 25 %[4] of the material comes from early middle English, between 1100 and 1300, and 75 % from Late Middle English, from between 1300 and 1500. The dominance of late Middle English texts reflects some important political and societal developments during the Middle English period. The Norman conquest of 1066 had a considerable impact on the nature of written communication on the British Isles. During the 12th-14th centuries, there were considerably fewer texts written in English. The English language lost in importance in comparison to Latin and Anglo-Norman French (Horobin & Smith, 2002). In the 15th century, English once again gained in importance as the strong connection to France was severed. These societal changes are thus also reflected in the composition of the corpus. When the diachronic development is discussed in the dissertation, particularly in Chapter 5, it is relevant to note that the early Middle English period is less well represented than the other periods of the corpora.

[3]In the Parsed Corpus of Early English Correspondence (Taylor et al., 2006), there are 383,822 words of correspondence from late Middle English.

[4]This proportion is based on the number of clauses (IPs in the corpus) coded as being part of texts from the early Middle English period.

The Middle English period is a period of great change and variation, which is also reflected in the corpus. Nominal and verbal morphology disappears more and more, which can be seen in the corpus annotation. In the Middle English corpus, case, nominative, accusative, genitive and dative, is no longer marked; instead, nominals are tagged for grammatical function, such as subject, object and second object.

3.1.3 The Early Modern English corpus

The Penn-Helsinki Parsed Corpus of Early Modern English is the largest of the historical corpora with its 1.7 million words. These words are distributed over 448 texts in the genres showed in Table 3.4.

Table 3.4: Texts in the Early Modern corpus

genre	No of texts	No of words	%
Proceedings, trials	16	137,249	7.9
Bible	12	133,585	7.7
Diary, private	21	127,689	7.3
Travelogue	19	122,145	7.0
Letters, private	129	116,423	6.7
Law	20	115,621	6.7
Fiction	18	112,438	6.5
Educational treatise	20	110,349	6.3
Drama, comedy	18	110,078	6.3
Handbook, other	19	105,435	6.1
History	18	103,769	6.0
Sermon	22	93,932	5.4
Philosophy	9	83,208	4.8
Science, other	13	77,446	4.5
Letters, non-private	71	60,771	3.5
Biography, other	9	50,490	2.9
Science, medicine	7	40,789	2.3
Biography, autobiography	7	36,436	2.1
total	448	1,737,853	100

Table 3.4 shows that there is a relatively even distribution between the genres. Unlike the corpora of Old and Middle English, there is no dominance of texts with a religious theme and the number of texts that constitute translations is much lower. Another difference in comparison to the Old and Middle English corpora is that the Early Modern English corpus

includes private diaries and letters that give records of a language that is closer to the spoken language. The Early Modern English corpus also contains proceedings from trials that also constitutes a window into the spoken language of the period.

Considering possible differences between informal and formal text genres, the fact that the genres letters, diaries and drama are included only from the Early Modern period onwards is important in terms of the historical development. If we would have had access to the same text genres for all periods, the development might have looked slightly different. The discussion about text genres is relevant also for the Late Modern English corpus, which is discussed in the next section.

3.1.4 The Late Modern British English corpus

The Penn Parsed Corpus of Modern British English is the last of the historical corpora used in this study. The period it covers stretches between about 1710 and 1914. It is slightly smaller than the other historical corpora with just under a million words. The genre distribution is about the same as the Early Modern English corpus, but instead of 448 different texts it contains 101 different texts.

Table 3.5 shows the genre distribution for the texts of the Late Modern British English corpus. As can be seen, there are several similarities between the Early and Late Modern English corpora. Both include a considerable proportion of correspondence, diaries and trial proceedings. The same concerns about the comparison between the Early Modern English Corpus and the earlier corpora also applies to the Late Modern English corpus.

3.1.5 The Present-Day English Corpus

The corpus used for the present investigation with respect to Present-Day English is the British National Corpus (BNC)[5], which is a corpus of present-day British English tagged for parts-of-speech. Two things sets the BNC apart from the previously presented corpora. Firstly, with around 100 million words, the BNC is considerably larger than the corpora of historical English. Secondly, the BNC contains transcribed spoken

[5]The material of the BNC was collected in the 1980s and 1990s in a project led by Oxford University Press. The aim of the creation of the corpus was to produce a 'wide cross-section of British English from the later part of the 20th century, both spoken and written' (http://www.natcorp.ox.ac.uk/corpus/).

Table 3.5: Texts in the Late Modern British English corpus

genre	No of texts	No of words	%
Travelogue	7	71,145	7.5
Drama, comedy	7	70,338	7.4
Diary	7	69,584	7.3
Educational treatise	9	64,839	6.8
Letters, private	8	66,362	7.0
Fiction	7	65,626	6.9
Law	7	65,748	6.9
Handbook, other	7	63,557	6.7
History	7	61,621	6.5
Proceedings, trials	3	58,973	6.2
Sermon	6	54,711	5.8
Bible	5	52,909	5.6
Science, other	6	53,449	5.6
Letters, non-private	4	33,826	3.6
Biography, other	3	30,072	3.2
Biography, autobiography	3	25,880	2.7
Science, medicine	3	23,147	2.4
Philosophy	2	17,108	1.8
total	101	948,895	100

language. The spoken part of the corpus amounts to about 10 %. The overall composition of the corpus is given in Table 3.6.

Table 3.6: Material of the BNC

Text genre	No of texts	No of words	%
Spoken demographic	153	4,233,955	4.3
Spoken context-governed	755	6,175,896	6.3
Written books and periodicals	2,685	79,238,146	80.6
Written-to-be-spoken	35	1,278,618	1.3
Written miscellaneous	421	7,437,168	7.6
total	4,049	98,363,783	100

Considering its size, the inclusion of spoken language, and the focus on informal language, the BNC is not really comparable to any of the other corpora presented here. The BNC has been used solely for a small study

on one particular type of clausal subjects in Present-Day English, namely *whether*-clauses. Out of all instances of the word *whether* in the corpus, a sample of 1,000 instances has been extracted, which forms the basis of the investigation of the alternation between the preposed clausal subject construction and the *it*+subclause construction in Present-day English in Chapter 7. The sample of 1,000 instances of the word *whether* from the BNC will henceforth be known as the BNC sample. The choice to focus on *whether*-clauses rather than including all clause-types is based on that fact that the word *whether* seems to be less ambiguous than for instance the word *that*, resulting in a sample with less irrelevant material. The word *whether* solely occurs in subordinate content clauses, while for instance the word *that* also acts as a personal or relative pronoun. Since the BNC sample was manually coded, it was important to keep down the proportion of irrelevant material.

After this brief introduction of the corpora used in the dissertation, the next section gives more information on how the corpora have been made use of to investigate the phenomena under discussion.

3.2 Method

As introduced above, the present investigation is based on data from a number of corpora of historical and Present-day English. Given that the object of investigation constitutes predicates that alternatively take a clausal subject and a subject *it* in conjunction with a subclause, different corpus-linguistic methods have been used to extract relevant sentences from the corpora. These methods differ between the historical corpora and the BNC.

The method used with respect to the historical corpora can be divided into three steps. First, coding queries were devised, using the CorpusSearch program to extract and code the relevant constructions and factors. The coding queries used in the dissertation are listed in the appendix. The second step was to extract a file containing only the code, which was then imported into the statistics software R (https://www.r-project.org/), where the data was analysed further.

With respect to the BNC, the web interface at http://corpus.byu.edu/bnc/ has been used to make coding queries for the small subsidiary investigation based on the BNC. A search for all instances of the word *whether* was made. Out of the result of this search, a sample of 1,000 instances of the word *whether* was extracted. This sample, the BNC

sample, was then coded manually with respect to the grammatical function of the *whether*-clause, weight/complexity values and information structure.

Below, the methodology used with respect to the historical corpora, whose analysis constitute the greater part of the dissertation, is described in more detail. In Section 3.2.1, there is a brief discussion on the measures of relative frequency in the corpora, which is relevant to the analysis of the data found in the corpora. In Section 3.2.2, the corpus annotation is presented and discussed. It is shown what the relevant syntactic phenomena of the dissertation look like in the corpora and what possibilities there are to search for them. This subsection also includes a discussion of the search program and the choices made with respect to particular coding queries.

3.2.1 Relative frequency

In the analysis of the corpus material, two different measures of relative frequency have been used, one for the BNC and one for the historical corpora.

When it comes to the BNC sample, relative frequency is based on number of instances per 1,000 instances of the word *whether*. With respect to the Penn Corpora of Historical English, relative frequency is calculated per 100,000 IPs. The IP-annotation in the corpora is used to represent a finite or non-finite clause. We thus have a measure of the relative frequency in relation to clauses rather than words. This has a number of advantages. First, the phenomena investigated in this dissertation are clause-level phenomena. If we want to determine how common preposed clausal subjects are, the most relevant measure is in terms of how many clauses contain preposed clausal subjects. Say that we have a corpus of 1000 words, which consists of 100 clauses. In this corpus, there are 10 instances of the preposed clausal subject construction. We thus have a relative frequency of one per 10 clauses or one per 100 words. Consider then a different corpus that also contains 1,000 words. In this second corpus, however, there are only 50 clauses. If the number of preposed clausal subject constructions is the same, we have a relative frequency of 0.5 instances per 10 clauses and one per 100 words. If we would like to compare these two corpora, we get two different results. In terms of instances per 10 clauses, there are half as many instances in the second corpus as in the first. In terms of instances per 100 words, we get the exact same relative frequency value. Of course, preposed clausal subjects can only occur once per clause. In order to represent correctly how common preposed clausal subjects are in the two corpora described, relative frequency in terms

of instances per 100,000 clauses thus gives a more accurate value. The reason that a similar measure of relative frequency has not been used in relation to the BNC is that there is no annotation available for either clauses or finite verbs (http://corpus.byu.edu/bnc/).

Further support for the adoption of a relative weight value based on IPs comes from the relation between number of IPs and the number of words in the corpora. Consider Table 3.7.

Table 3.7: Absolute and relative frequencies of IPs in the corpora

periods	tokens	per million words
OE	236,046	157,364
ME	217,103	180,919
EME	150,804	88,708
LME	127,337	134,039

As can be seen in the table, there are considerable differences between the corpora when it comes to the number of IPs per one million words. Strikingly, the Early Modern English (EME) corpus contains a considerably lower number of IPs per million words in comparison to the other corpora. Considering what has been said above about the difference between calculating relative frequency in number of clauses or number of words, the use of relative frequency per 100,000 clauses seems to be highly motivated for the Penn Corpora of Historical English.

After having considered the measure of relative frequency, in the next section we proceed to a more detailed account of the annotation of the Penn Corpora of Historical English and the York-Toronto-Helsinki Parsed Corpus of Old English Prose.

3.2.2 Annotation

In order to find the relevant constructions, we need to know how they are marked up the corpora. The corpora, except for the BNC, are syntactically annotated for phrasal and functional categories. In this section, the annotation in the corpora in relation to the constructions under investigation is presented and exemplified.

According to Taylor et al. (2003); Kroch & Taylor (2000); Kroch et al. (2005, 2010), the main motivation for annotation is to facilitate automated searches rather than to give a linguistically correct representation of the sentences in the texts. Furthermore, to avoid subjective

judgements, certain ambiguous categories are lumped together. For example, no distinction is made between adjectival and verbal participles or between arguments and adjuncts. VPs, whose boundaries sometimes are indeterminate, are also not annotated. The attachment of adjunct phrases frequently give rise to ambiguous structures. In the corpora, this problem has been dealt with by annotating adjunct phrases as attached as high as possible, whenever there is ambiguity. This means that the sentence *I saw the man with the telescope* always will be given the analysis in which the PP *with the telescope* modifies the verb *see*. As this choice affects constituent structure, it is relevant with respect to the design of coding queries in the investigation.

In order to initiate the discussion about the nature of the annotation used in the historical corpora, the sentence in (1) is given as a point of departure. The sentence is an example of the *it*+subclause construction.

(1) a. At the University, it is optional to pursue Classics. (BAIN-1878,380.312)

b. Annotated example:
((IP-MAT
(PP (P At) (NP (D the) (N University))) (, ,)
(NP-SBJ-1 (PRO it))
(BEP is)
(ADJP (ADJ optional))
(IP-INF-1 (TO to) (VB pursue) (NP-OB1 (NS Classics))) (. .))
(ID BAIN-1878,380.312))

In (1-a), we first see the sentence without annotation. In (1-b), the annotated example is given, as it comes out in the corpus file. Notice first the round brackets, which are used to represent phrase structure. Each constituent is given a label for syntactic or phrase category. From the outside in, we have an IP (inflection phrase), which is furthermore given the functional label -MAT for matrix. All IPs that are outermost in the phrase structure, i.e. are non-embedded, are given the label IP-MAT for matrix clause. The IP-MAT here dominates five constituents: (i) the PP *At the university*, (ii) the NP-SBJ (NP subject) *it*, (iii) the 3rd person singular form of the verb *be*, *is*, (iv) the ADJP (adjective phrase) *optional*, and, finally, (v) the IP-INF (inflection phrase with the functional label infinitive) *to pursue Classics*. Below the phrase structure annotation of the sentence, we have a line giving an identification tag for this particular sentence (ID BAIN-1878,380.312), which says that the sentence comes from *Education as a science* by Alexander Bain, published in 1878. The

last numbers of the identification tag (380.312) tells us the place in the text where this particular sentence occurs.

One feature of the tagging that has not been mentioned is the indexation -1 given to the NP subject and the IP infinitival clause. This type of annotation will be further discussed in the next section. In short, it is a way of indicating a coreference relation between these two constituents.

In the following, the annotation of the preposed clausal subject construction and the *it*+subclause construction, respectively, are presented and discussed.

The clausal subject construction

All clauses in the corpus are given the annotation IP. They are furthermore specified for their function. The annotation IP-MAT is used for matrix clauses, while the annotation IP-SUB is used for subclauses. IP-INF is the annotation for infinitival clauses. The IP-SUBs are contained within different CPs[6] (complementizer phrases) depending on their type. Examples of types of subordinate clauses include SBJ (subject), THT (*that*-clause), ADV (adjunct clause), REL (relative clauses), QUE (questions), FRL (free relative clause). The relevant categories in the search for finite and non-finite clausal subjects are CP-THT, CP-QUE and IP-INF. With respect to *that*-clauses, *wh*-clauses, and infinitival clauses tagged as subjects, these have the extension -SBJ. An example of how a preposed clausal subject is annotated is given in (2).

(2) a. And, sire, that ther hath been many a good womman, may lightly be preved. (CMCTMELI,220.C2.134)
'And, sir, that there has been many a good woman, may easily be proved.'

b. Annotated example:
((IP-MAT-SPE
(CONJ And) (, ,)))
(NP-VOC (N sire)) (, ,)))
(CP-THT-SBJ-SPE (C that)
(IP-SUB-SPE (NP-SBJ-1 (EX ther))
(HVP hath)
(BEN been)
(NP-1 (Q many) (D a) (ADJ good) (N womman)))) (, ,)))
(MD may)
(ADVP (ADV lightly))

[6]The CP annotation is present for all finite subordinate clauses in the corpus.

(BE be)
(VAN preved) (. .)))
(CMCTMELI,220.C2.134))

The *that*-clause *that ther hath been many a good womman* is here classified as CP-THT-SBJ-SPE. The SPE annotation signifies that the sentence constitutes direct speech.

As will be seen, not all subclauses marked as subjects occur in a clause-initial position. They also occur in subject positions immediately after the finite verb in subject-aux inversion or immediately following a conjunction in a subclause. Subclauses that occur in a clause-final position are, however, never marked as subjects. Consider the sentence given in (3), which represents what I call the null+subclause construction.

(3) a. Thanne is toold how Salomon byldide the temple of Jerusalem, and an hous to himself.
'Then it is told of how Salomon built the temple of Jerusalem, as well as a house for himself.'
(CMPURVEY,I,21.1005)

b. Annotated example:
((IP-MAT
(NP-SBJ-1 *exp*)
(ADVP-TMP (ADV Thanne))
(BEP is)
(VAN toold)
(CP-QUE-1 (WADVP-2 (WADV how))
(C 0)
(IP-SUB (ADVP *T*-2)
(NP-SBJ (NPR Salomon))
(VBD byldide)
(NP-OB1 (NP (D the) (N temple)
(PP (P of) (NP (NPR Jerusalem)))) (, ,)
(CONJP (CONJ and)
(NP (D an) (N hous)
(PP (P to) (NP (PRO+N himself)))))))) (. .)))
(CMPURVEY,I,21.1005)

In (3), there is a clause-final subclause, which constitutes the only argument of the passive predicate *be told*. This subclause is not marked out as a subject in the corpus. Instead, an empty subject has been added in the annotation of the clause, which has the form *exp* (for expletive).

In sentences where there is no subject in a clause-initial or clause-final subject position and a subject is required, an empty subject is inserted in the annotation. Apart from the category *exp*, there are three other categories of empty subjects: (i) subjects elided under conjunction, (NP-SBJ *con*), (ii) arbitrary subject in ECM infinitives, (NP-SBJ *arb*), and other empty subjects, (NP-SBJ *pro*).

The *it+subclause construction*

The pronoun *it* in the corpora is simply represented as the word *it*[7] (or one of the forms *hit*, *hit*, *hitt*, *hytt*, *yt*, *ytt* or *itt*) dominated by the category PRO (for pronoun). An important distinction made in the dissertation is the one between thematic and non-thematic *it*, i.e. whether or not *it* is associated with a thematic role. Given the corpus annotation, there is no immediate way to distinguish between thematic and non-thematic *it*. An example of the way a pronoun *it* is represented in the corpora is given in (4) for a weather-verb sentence.

(4) a. it is even drizzling a little
(ID CARLYLE-1837,1,138.38)
b. Annotated example:
((IP-MAT
(NP-SBJ (PRO it))
(BEP is)
(FP even)
(VAG drizzling)
(NP-MSR (D a) (ADJ little)) (. .)
(ID CARLYLE-1837,1,138.38))

In (4), the subject *it* is simply represented as a pronoun (PRO) within a nominal phrase marked as subject (NP-SBJ).

When a subject *it* is accompanied by a subclause in the *it*+subclause construction, there is a coindexation between the subject and the subclause. Consider the sentence in (5).

(5) a. It is impossible to be exact in the Time.
(ID HOLMES-TRIAL-1749,20.327)
b. Annotated example:
((IP-MAT
(NP-SBJ-1 (PRO It))

[7] I make use of the word *it* as a cover term for this lexeme, including its spelling alternation.

(BEP is)
(ADJP (ADJ impossible))
(IP-INF-1 (TO to) (BE be)
(ADJP (ADJ exact)
(PP (P in) (NP (D the) (N Time))))) (. .))
(ID HOLMES-TRIAL-1749,20.327))

As can be seen in (5), the subject *it* is coindexed (*-1*) with the subclause *to be exact in the Time*. By making use of the query specification *sameIndex*, it is possible to label all sentences where this coindexation occurs.

When it is a demonstrative pronoun, rather than a subject *it* that occurs in conjunction with a subclause, a different representation in the corpus is given. In these cases, the subclause is given the tag -PRN- (parenthetical) and is coindexed with a trace marked *ICH* (= interpret constituent here) in the position next to the demonstrative pronoun. Consider the sentence in (6).

(6) a. This is my Commaundement, that ye loue one another, as I haue loued you.
(AUTHNEW-E2-P1,15,1J.395)
'This is my commandment, that you love one another, as I have loved you.'

b. Annotated example:
((IP-MAT
(NP-SBJ (D This) (CP-THT-PRN-SPE *ICH*-1))
(BEP is)
(NP-OB1 (PRO my) (N Commaundement)) (, ,)))
(CP-THT-PRN-SPE-1 (C that)
(IP-SUB-SPE (NP-SBJ (PRO ye))
(VBP loue)
(NP-OB1 (ONE one) (D+OTHER another)) (, ,)
(PP (P as)
(CP-ADV-SPE (WADVP-2 0)
(C 0)
(IP-SUB-SPE (NP-SBJ (ADVP *T*-2)
(PRO I))
(HVP haue)
(VBN loued)
(NP-OB1 (PRO you))))))) (. .)))
(AUTHNEW-E2-P1,15,1J.395))

In (6), the appositive *that*-clause *that ye loue one another, as I have loued you* is coindexed with a trace, marked with the same category (CP-THT-PRN-SPE *ICH*-1), next to the phrase which it modifies (the subject demonstrative). Structures such as the one in (6) do not form part of the present investigation and will not be further discussed.

3.2.3 Use of coding queries

After having described the annotation of the corpora, a few words need to be spent on how the annotation has been used to search for and code the relevant sentences. As will be recalled, the search program CorpusSearch 2 (Randall, 2005-2007) has been used to code the sentences in the corpora. As described in the introductory chapter, the three constructions we are considering in this dissertation are: (i) the preposed clausal subject construction, (ii) the *it*+subclause construction and (iii) the null+subclause construction. As a consequence of their central role in the dissertation, the coding query used to identify these constructions is here explained in more detail. Consider the coding query represented in (7), which is given the same form as it has in the coding file (.c).

(7) //Constructions:
1: {
nonextra: ((CP*SBJ* exists) OR (IP-INF-SBJ* exists)) AND (!CP-FRL-SBJ* exists)
extraexp: (NP-NOM*|NP-SBJ* iDoms PRO*) AND ((NP-NOM*|NP-SBJ* sameIndex IP-INF*|CP-THT*|CP-QUE*) OR (NP-NOM-x hasSister IP-INF-x|CP-THT-x|CP-QUE-x)) AND (NP-NOM*|NP-SBJ* hasSister !NP-MSR*|NP-2*|NP-1*|NP-NOM)
extranonexp: (NP-NOM*|NP-SBJ* iDoms exp) AND ((NP-NOM*|NP-SBJ* sameIndex IP-INF*|CP-THT*|CP-QUE*) OR (NP-NOM-x hasSister IP-INF-x|CP-THT-x|CP-QUE-x)) AND (NP-NOM*|NP-SBJ* hasSister !NP-MSR*|NP-2*|NP-1*|NP-NOM)
z: ELSE }

In (7), I have given the three constructions three different code labels, *nonextra*, *extraexp* and *extranonexp*. Let us look at these in turn, starting with the query for the preposed clausal subject construction. The relevant part of the query is repeated in (8).

(8) nonextra: ((CP*SBJ* exists) OR (IP-INF-SBJ* exists)) AND (!CP-FRL-SBJ* exists)

The query with the label *nonextra*, given in (8), calls for a CP tagged as subject or an IP-INF tagged as subject exists in the clause. Furthermore, it excludes from the result any free relative clauses tagged as subject that might exist in the corpora. The exclamation mark (!) is used to negate an argument of the coding query.

The query for the second construction, the *it*+subclause construction, is repeated as (9).

(9) extraexp: (NP-NOM*|NP-SBJ* iDoms PRO*) AND ((NP-NOM*|NP-SBJ* sameIndex IP-INF*|CP-THT*|CP-QUE*) OR (NP-NOM-x hasSister IP-INF-x|CP-THT-x|CP-QUE-x)) AND (NP-NOM*|NP-SBJ* hasSister !NP-MSR*|NP-2*|NP-1*|NP-NOM)

The query for the label *extraexp*, given in (9), calls for an NP tagged as having nominative case or being a subject that immediately dominates a pronoun (PRO*). It then specifies, using the command *sameIndex*, that this NP should have the same index as an infinitival clause (IP-INF*), a *that*-clause (CP-THT*) or a *wh*-clause (CP-QUE*).

The *sameIndex* command does not work for the Old English corpus (YCOE), where instead of index numbers, the coindexation is managed with the tag -x in the annotation. Therefore an alternative is introduced where an NP tagged as -NOM-x should have as its structural sister, *hasSister*, a clause which is also tagged as -x.

Finally, we have the query for the null+subclause construction, which is repeated as (10). The query here is almost the same as the one for the *it*+subclause construction. The difference between the two queries concerns the element *exp*, given in bold face.

(10) extranonexp: (NP-NOM*|NP-SBJ* iDoms ***exp***) AND ((NP-NOM*|NP-SBJ* sameIndex IP-INF*|CP-THT*|CP-QUE*) OR (NP-NOM-x hasSister IP-INF-x|CP-THT-x|CP-QUE-x)) AND (NP-NOM*|NP-SBJ* hasSister !NP-MSR*|NP-2*|NP-1*|NP-NOM)

The query with the label *extranonexp*, given in (10), which represents the *nullsubject+subclause* construction, calls for the same structures as the query for *extraexp*, except for the fact that instead of the nominative subject NP immediately dominating a pronoun, now the subject NP is instructed to immediately dominate an *exp*, an empty subject.

In the above, the most important coding query, the one for the three constructions under discussion in the dissertation, has been presented. As mentioned, a list of all coding queries used in the dissertation is given in the appendix, Section A2.

3.3 Summary

In the present chapter, the material and method used in the dissertation have been presented and discussed. The investigation is corpus-based and makes use of corpora of both historical and Present-day English. The historical corpora consist of the the Penn Corpora of Historical English (Kroch et al., 2000, 2005, 2010) and the *The York-Toronto-Helsinki Parsed Corpus of Old English Prose* (Taylor et al., 2003). For Present-day English, the material derives from the British National Corpus (BNC). In terms of the possibility to make comparisons between the different historical corpora, some potential issues were discussed. For one thing, the two later corpora of Early and Late Modern English contain a number of text genres not represented by the earlier corpora. In particular, they include personal diaries and letters that can be assumed to represent a language that is less formal and possibly closer to the spoken vernacular. Apart from giving a presentation of the corpora, the chapter also includes a description of the annotation of the corpora. The annotation associated with the three constructions under discussion, the preposed clausal subject construction, the *it*+subclause construction and the null+subclause construction, have been discussed in more detail. It was shown how these two constructions can be found and coded in the corpora.

Part II

Syntax and argument structure

Chapter 4

Background

In the first part of the dissertation, the theoretical tools and methodology were presented in conjunction with the corpora forming the material investigated. In this part, I present a study of the syntax and argument structure of the alternations established in the introduction, i.e. the alternation between preposed clausal subjects, the *it*+subclause construction and the null+subclause construction. The present chapter gives a background to the questions investigated and a summary of previous research. In Chapter 5, my own analysis of the argument structure and syntax of the alternations is given.

The structure of the chapter is as follows. In the first section, studies on clausal subjects in Present-day English are discussed, followed by studies on the *it*+subclause construction in the second section. The third section concerns clausal subjects and the *it*+subclause construction in historical English.

4.1 Clausal subjects in PDE

Within the generative syntax literature, the question of whether subordinate clauses can constitute subjects in Present-day English have attracted considerable attention. As a point of departure, let us take Koster (1978). His study relates to a discussion that arose in the 1960s and 1970s on the status of clausal subjects within transformational grammar (e.g. Rosenbaum, 1967; Emonds, 1976; Koster, 1978; Stowell, 1981). The question concerns whether the subclause (within square brackets) in a sentence such as (1) constitutes the subject of the clause in which it occurs. The notion of subject in this discussion of subclauses in transformational grammar is to be understood as a structural subject, i.e. a phrase occurring in

a particular position in the constitutent structure which is associated with subjecthood. The definition of subjecthood thus departs from the one given in Section 2.3.1, where both syntactic and morphosyntactic properties of subjects are taken into account.

(1) [That the doctor came] surprised me. (Koster, 1978: 53)

In (1), the *that*-clause *that the doctor came* occurs in a clause-initial position followed by the verbal predicate *surprised* and the object *me*. By simply looking at this individual sentence, it is not possible to decide the structural position of the subclause. However, based on grammaticality judgements on the alternations shown in (2), Koster (1978) concludes that *that*-clauses cannot constitute structural subjects. The alternations in (2) show three environments where structural subjects occur: (i) within a subclause between the subordinating conjunction and the finite verb, (ii) immediately following the finite verb in a question, and (iii) between a fronted constituent and the finite verb.

(2) a. subordinate clause:
   (i) Although it may depress you [that the house is empty], it pleases me.
   (ii) *Although [that the house is empty] may depress you, it pleases me.
   b. subject-auxiliary inversion:
   (i) Did it please you [that John showed up]?
   (ii) *Did [that John showed up] please you?
   c. fronting:
   (i) Such things, it doesn't prove.
   (ii) *Such things [that he reads so much] doesn't prove.

According to Koster, the examples in (2-a-ii), (2-b-ii) and (2-c-ii) are ungrammatical, while the constructions in (2-a-i), (2-b-i), (2-c-i), which have pronominal subjects, are grammatical. These judgements give support to the hypothesis that subclauses cannot constitute subjects. The ungrammaticality of the preposed clausal subject constructions in (2) is not undisputed. For example, there are studies claiming that the sentences marked as ungrammatical in (2) are not ungrammatical but unacceptable for other reasons (e.g. Delahunty, 1983; Miller, 2001; Davies & Dubinsky, 2009). According to these studies, given the right conditions, the sentences in (2-a-ii), (2-b-ii) and (2-c-ii) could be judged as grammatical. Consider the sentences in (3), given by Davies & Dubinsky (2009), Delahunty (1983) and Miller (2001: 697), respectively.

(3) a. Is [that I am done with this homework] really amazing?
b. Who does [that the world is ending] upset so terribly that they have decided to abandon the planet?
c. Descartes claimed that the two lines in figure C were parallel and provided a proof based on his second theorem. This proof was in fact mistaken. From his first theorem on the other hand, [that the two lines are parallel] certainly does follow, but remarkably, Descartes apparently never noticed this.

Davies & Dubinsky (2009), Delahunty (1983) and Miller (2001) argue that the environments used by Koster (1978) also can be occupied by *that*-clauses. In (3-a), we see Davies & Dubinsky's (2009) example of a *that*-clause occurring in the subject position in subject-auxiliary inversion, in (3-b) we see Delahunty's (1983) example of inversion with a *wh*-phrase, and, in (3-c), we see Miller's (2001) example of a *that*-clause, also occurring in subject position, following the clause-initial topicalised phrase *from this first theorem*. It should be noted that Miller's example is taken from naturally occurring discourse, while Davies and Dubinsky's and Delahunty's examples are constructed.

As pointed out above, the test environments shown in (2) concern structural subjecthood, i.e. whether an element occurs in the subject position, Spec,IP in English. For the purposes of my investigation, functional subject properties are also relevant. Davies & Dubinsky (2009) uses the test environments in (4) to argue that *that*-clauses can constitute morphosyntactic subjects. The test environments are obligatory raising, verb agreement, licensing of the adverb *equally*, and hosting emphatic reflexives.

(4) a. subject raising:
[That Shelby lost it] appears to be true.
b. verb agreement:
[That the march should go ahead and that it should be canceled] have been argued by the same people at different times.
c. licensing of the adverb *equally*:
[That he'll resign and that he'll stay in office] seem at this point equally possible. (McCloskey 1991: 564)
d. hosting emphatic reflexives:
[That there were twenty-five miles to go] was itself enough to discourage Edwin.

In (4-a), we see a subordinate clause raised to the subject position of the verb *appear*. In (4-b), the two coordinated clause-initial *that*-clauses seem

to trigger plural morphology on the auxiliary *have*. The examples in (4-c) and (4-d) show that the clausal subject construction can occur with the adverb *equally* and with emphatic reflexives.

Different analyses of clausal subjects have beeen made in recent literature in relation to the type of data shown so far in this chapter. Within the transformational tradition, two examples are Alrenga (2005) and Davies & Dubinsky (2009). Within LFG, Bresnan (2001) should be mentioned. Recall from Chapter 2 that a distinction is made within LFG between the subject position in the constituent phrase structure, c-structure, and the grammatical subject function in the f-structure, i.e. between structural and functional subjecthood. Subclauses can be analysed as subjects even though they do not occur in the subject position in English. Consider the sentence in (5), from Bresnan (2001: 20).

(5) [That languages are learnable] is captured by this theory.

Bresnan's (2001: 20-21) analysis of the sentence in (5), which contains the clausal subject *that languages are learnable* can be represented as in Figure 4.1 (see top of next page). As can be seen in the figure[1], the CP *that languages are learnable* occurs in a fronted position, adjoined to the IP, rather than in the subject position, Spec,IP. However, even though it occurs in a fronted TOP position, the *that*-clause is still identified as the functional subject, SUBJ, in the f-structure. This relation can be represented by the functional equation (↑ TOP) = (↑ SUBJ), which in Figure 4.1 is represented by a line between TOP and SUBJ in the f-structure. The equation of the subject and topic in this sentence follows from the principles of LFG. Since the verb *capture* requires a subject, the completeness principle (i.e. the principle which says that the f-structure must contain the arguments required by the predicate) states that a subject must be identified. The so-called grammaticalised discourse function TOP cannot stand on its own, but must be identified with some other grammatical function. The two possible choices are the SUBJ or the OBL$_\theta$. Since the OBL$_\theta$, as can be seen in its f-structure, is already associated with its own PRED value, the only available function for the TOP to be functionally equated with is the SUBJ.

Summarising the discussion above, it can be said that there is conflicting evidence on the subject status of subordinate clauses. The sentences

[1] Another thing that should be mentioned about the f-structure in 4.1 is the fact that the copula *be* does not have its own PRED-value. The copula simply provides information about tense, aspect and modality (Bresnan et al., 2016: 110). For a discussion about different ways to analyse copulas within LFG, see Dalrymple et al. (2004), Nordlinger & Sadler (2006) and Attia (2008).

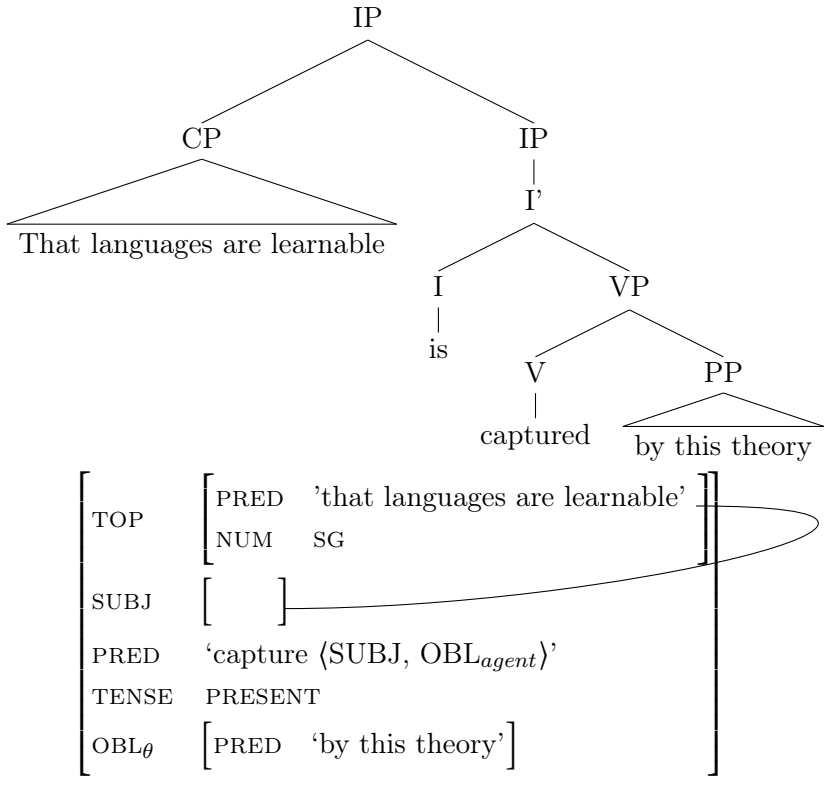

Figure 4.1: C-structure and f-structure for the sentence *That languages are learnable is captured by this theory.*

presented in (2) are given as support for the hypothesis that subclauses cannot occur in the structural subject position. However, the sentences in (3), assuming that they are acceptable, provides counterevidence to such a hypothesis. The sentences in (3) suggest that subclauses can occur in the structural subject position, and that the sentences in (2) are unacceptable for other reasons. In Chapter 5, it will be shown that there is some support for the hypothesis that infinitival clauses sometimes occur in a structural subject position in Early and Late Modern English, while there is no support for the hypothesis that *that*-clauses and *wh*-clauses occur in this position during these periods. With respect to functional subjecthood, the behaviour of the *that*-clauses in (4) supports an analysis of *that*-clauses as functional subjects. In Chapter 5, evidence will be given from Early and Late Modern English that also *wh*-clauses and infinitival clauses function as morphosyntactic subjects.

4.2 The *it*+subclause construction in PDE

The preposed clausal subject construction discussed in the preceding section alternates with the *it*+subclause construction. Three examples of this latter construction from the historical corpora are given in (6). The sentence in (6-a) derives from Early Modern English, while the sentences in (6-b) and (6-c) derive from Late Modern English.

(6) a. It is sayde [that I wuld haue saued the senators].
(BOETHCO-E1-P1,20.47)
b. It seems [that there are consulting physicians in Africa].
(READE-1863,212.260)
c. It is good for you [to make him well].
(READE-1863,212.264)

The sentences in (6) all include a subject *it* in conjunction with a propositional subclause. In (6-a), the *it*+subclause construction occurs in conjunction with a passive construction, in (6-b), with an impersonal verb, and, in (6-c), with a copular construction.

In this section, a number of studies concerning the analysis of the construction types in (6) are presented, among them Seppänen et al. (1990); Seppänen & Herriman (2002); Kaltenböck (1999, 2005); Shahar (2008). In addition, reference is also made to Berman (2003), who discusses the same constructions for Present-day High German within an LFG framework. Two questions about the construction types in (6) concern, firstly, the status of the subject *it*, whether it is thematic or nonthematic, and, secondly, the status of the subclause, whether it constitutes an adjunct, a complement or a subject. These two issues are dealt with in the next two subsections.

4.2.1 Subclause as complement or adjunct

With respect to the analysis of the status of the propositional subclause as a complement or an adjunct, different constituency tests can be applied. In this section, three tests will be discussed: (i) VP-topicalisation, (ii) *wh*-movement with pied-piping, and (iii) extractions.

To begin with, we may note that Shahar (2008) finds support both for complement and adjunct status of the subclause in the *it*+subclause construction in English. First, let us consider VP-topicalisation, which is used by Shahar (2008) as a test of the status of the subclause. If VP-topicalisation is possible, this suggests that the predicate and the

propositional subclause form a constituent together and that the subclause should be analysed as a complement. Consider the sentences in (7).

(7) a. %They wondered whether it was important [that you call me back], and important [that you call me back], it was.
(Shahar, 2008: 32)
b. %They wondered whether it seemed [that John likes Mary], and seem [that John likes Mary], it did.
(Shahar, 2008: 32)

For the monadic predicates in (7-a) and (7-b), where it is tested whether the predicate and the *that*-clause can be fronted together, the judgements are divided. For some speakers, the verb forms a constituent together with the clausal argument, and for other speakers it doesn't.

Seppänen (1986) also discusses VP-topicalisation in relation to the *it*+subclause construction. Seppänen draws a distinction between structures where the subclause is an argument to the verbs *seem* or *appear* and the cases where the subclause is an argument of a predicate adjective, as in Shahar's example in (7-b) above. Consider the examples in (8) and (9)

(8) Margaret feared that it would then be (seem, appear) obvious [that we were wrong],
a. *and be (seem, appear) obvious [that we were wrong] it would.
b. ?and be (seem, appear) obvious it would [that we were wrong].
c. and be (seem, appear) obvious it would.

(9) Margaret feared that it would then seem (appear) [that we were wrong],
a. and seem (appear) [that we were wrong] it would.
b. *and seem (appear) it would [that we were wrong].
c. and seem (appear) so it would.

For the examples in (8) and (9), Seppänen argues that there is a difference between the predicates *seem/appear obvious* and *seem/appear* when it comes to the position of the clausal argument. The predicate *seem/appear obvious* cannot be fronted together with the clausal argument, while this appears to be possible with the predicate *seem/appear*.

Seppänen (1986) and Shahar (2008) differ in the judgements they have for the sentences where *seem* forms a constituent together with the subclause. Seppänen accepts it as grammatical, while Shahar reports of both grammaticality and ungrammaticality with different speakers. It is necessary to note here that the constructed examples of VP-topicalisation

are problematic in the sense that it is hard to imagine a context where they would be natural, which might affect the judgements[2]. Furthermore, VP-topicalisation in conjunction with subclauses involves center-embedding[3], which is commonly seen as dispreferred with respect to processing (cf. Karlsson, 2007).

Shahar also uses *wh*-movement of the predicate with pied-piping as a test for the status of the propositional subclause. According to Shahar (2008: 35), it shows that the subclause consistently does not form a constitutent together with the predicate. Three of Shahar's (2008: 36) sentences are given in (10), where the *wh*-word together with its dependent occur as a constituent.

(10) a. *If he's kissing her, how likely that John loves Mary would it be?
b. *If he's kissing her, how clear that John loves Mary would it be?
c. *If he's kissing her, how important that Mary isn't John's wife would it be?

The judgements reported by Shahar (2008: 36) for the sentences in (10) suggest that the predicate and the subclause do not form a constitutent, for any of the predicates shown. However, to a certain extent, *wh*-movement of the predicate with pied-piping suffers from the same shortcomings as VP-topicalisation as a test in the sense that it also involves center-embedding.

Another test is extraction, which does not involve center-embedding and might thus be better as a test. Here, Shahar (2008: 38) reports an interesting difference between predicates, specifically when it comes to the extraction of an adjunct out of the subclause. Extraction of an adjunct out of an adjunct is typically seen as impossible, while extraction of an adjunct out of a complement is possible. The difference between predicates mentioned by Shahar will be discussed again in Chapter 5. Consider the sentences in (11), where the adjunct *how*-phrase is supposed to be dependent upon the verb *kiss* within the *that*-clause.

[2]Shahar (2008: 35) notes that both VP-topicalisation and *wh*-movement with pied-piping are 'rarely used by most speakers', implicitly indicating that this might be important for the judgements given.

[3]According to Karlsson (2007: 367), center-embedding involves subordinate clauses that 'have words of the superordinate clause both to their left (excluding subordinators and coordinators) and to their right'.

(11) a. ?How is it possible [that John kissed Mary]? (With passion)
b. ?How is it clear [that John kissed Mary]? (With passion)
c. How is it likely [that John kissed Mary]? (With passion)
d. How does it seem [that John kissed Mary]? (With passion)

In (11), the two raising predicates *be likely* and *seem* in (11-c) and (11-d) allow extraction of the adjunct *how*-phrase, while the adjectival predicates in (11-a) and (11-b) do not allow extraction of an adjunct out of their subclause. The judgements in (11) suggest that the subclause in (11-a) and (11-b) is an adjunct, while the subclause in (11-c) and (11-d) is a complement. As will be seen in the next chapter, this is also the analysis adhered to in this dissertation.

4.2.2 Thematic or nonthematic subject *it*

As mentioned in the introduction, the analysis of the subclause as complement or adjunct is also connected to the analysis of the subject *it* as thematic or nonthematic. As will be seen in the next chapter, for a consistent argument structure, the analysis of the subclause as a complement leads to the analysis of the subject *it* as nonthematic; conversely the analysis of the subclause as an adjunct leads to the analysis of the subject *it* as thematic. However, there are also independent tests for the evaluation of the thematicity of the subject. One such test involves using control constructions. There is assumed to be a difference between argumental pronouns and nonargumental expletives (cf. Chomsky, 1981) in whether they can control an 'empty' subject in a control construction. Consider the difference between the sentences in (12).

(12) a. Before snowing, it often rains.
b. *There is often a party here right before being a wake.

In (12-a), the pronoun *it* controls (i.e. is functionally equated to) the subject of the non-finite clause *before snowing*. In (12-b), the subject *there* in the main clause does not seem to be able to license the lack of an overt subject in the nonfinite clause, *before being a wake*. This difference supports the conclusion that *it* is argumental and *there* nonargumental.

Shahar (2008: 29) claims that the subject *it* in the *it*+subclause constructions in (13) can control the subject of the fronted nonfinite clause[4].

[4]In fact, Shahar (2008: 29) claims that the subject *it* and the subclause *to smoke in a bar* function as a controller together.

(13) a. Besides being illegal in NYC, it is unhealthy [to smoke in a bar].
b. Before seeming/appearing likely, it was unlikely [that John would get the job].

In (13-a), it is possible to leave out the subject of the nonfinite clause *being illegal in NYC*, which is equated with the subject of the matrix clause. In (13-b), it is likewise possible, according to Shahar, to leave out the subject of the phrase *seeming/appearing likely*.

Shahar (2008: 30) takes the grammaticality of the structures in (13) as evidence that *it* is theta-role-bearing. However, this conclusion is based on the assumption that there is a two-way distinction between non-theta-role-bearing expletives and theta-role-bearing arguments[5]. In this dissertation, I instead make the assumption that there is one distinction to be made in terms of argument status (lexical selection) and one distinction to be made in terms of thematicity (whether the phrase is associated with a thematic role). The subject *it* in for example (13-b) might be argumental, but still nonthematic. As will be seen in the next chapter, this is the analysis subscribed to in this dissertation.

A second test to evaluate the thematicity of the subject *it* is whether it can be deleted in coordination with thematic or nonthematic subjects.

Kaltenböck (1999: 57) gives the example in (14), intended to show that a subject *it* in the *it*+subclause construction can be deleted in coordination with what is claimed to be a clearly referential subject *it*.

(14) A: Have you heard of the bank robbery in London?
B: Yes, *it's* terrible, and (*it's*) hard [to believe that they actually got away with it].

Assuming that the *it* of *it's terrible* refers back to the bank robbery in the question or to the situation in general, thus being thematic, and that second conjunct subject deletion requires the subjects to match in thematicity, the example in (14) supports the conclusion that the *it* in the *it*+subclause construction in (14) is thematic.

Although the judgement for the sentence in (14) suggests that the subject *it* is thematic, the test was only applied to one type of predicate, an adjectival predicate. It is possible second conjunct subject deletion in

[5]Chomsky (1981: 325) makes a distinction between theta roles, concerning the syntactic argument structure of the predicate, and thematic relations, concerning the thematic relation of an argument in relation to an action or event. The *it* in weather-verb constructions such as (12-a) is claimed to take the theta role *quasiargument*, which does not express a thematic relation.

connection with raising verbs such as *seem* might work differently. This also appears to be the case. Consider the constructed dialogue in (11-b).

(15) A: Have you heard of the bank robbery in London?
B: Yes, *it's* terrible, and ?(*it*) seems [that they actually got away with it].

In (15), it is doubtful whether the subject *it* can be left out in the second conjunct in connection with the verb *seem*. If the subject *it* cannot be left out in (15), this supports the hypothesis that the subject *it* is indeed thematic in conjunction with adjectival predicates, but nonthematic in conjunction with raising predicates. Such a distinction will be argued for in the Chapter 5.

The present section and the preceding section have been concerned with the status of the subject *it* and the subclause in the *it*+subclause construction. Certain tests have been identified to determine whether the subject *it* is thematic or nonthematic, and whether the subclause is a complement or an adjunct. In the next section, previous LFG analyses of the *it*+subclause construction are discussed.

4.2.3 Previous LFG analyses

When the *it*+subclause construction has been discussed within LFG, this has to my knowledge mainly been done in connection with the phenomenon of raising. The LFG analysis has similarities to the way in which the *it*+subclause construction has been analysed generally within generative and traditional grammar, where the subject *it* has been seen as a formal (place-holder) subject (e.g. Radford, 2004)[6]. In the Lexical Mapping Theory presented in Bresnan et al. (2016: 340), the argument structure and argument-to-function mapping of the raising verb *seem* are given as in (16).

[6]Radford (2004: 291) gives for instance the sentence in (i) and claims that the subject *it* is an expletive which '(by virtue of being non-referential) carries no interpretable θ-features'.

(i) It can be difficult to come to terms with long term illness.

The *it*+subclause construction in (i) does not contain a raising predicate. Nonetheless, the subject *it* is analysed as non-thematic by Radford.

(16) Argument structure for *seem*:

seem	_	⟨*experiencer*,	*proposition*⟩
	[–r]	[–o]	[–o]
	|	|	|
	SUBJ	OBL_{exp}	XCOMP/COMP

In (16), the verb *seem* takes three arguments, where the first argument constitutes an empty argument role that does not have a thematic role. Following the mapping principles given in Bresnan et al. (2016: 334), the empty argument role is mapped to SUBJ, the *experiencer* to OBL_{exp} and the *proposition* to XCOMP or COMP. This argument structure accounts, among other things, for the fact that the subject of the verb *seem* can be non-thematic *there* or *it*. An example of the verb *seem* taking a non-thematic subject *there* or *it* is given in (17).

(17) a. There seems to me [to be a problem with the proposal]. (Bresnan et al., 2016: 340)
b. It seems to me [that there is a problem with the proposal]. [constructed]

In (17-a), *there* is the subject of *seem* and also the subject within the XCOMP *to be a problem with the proposal*. The prepositional phrase *to me* is linked to OBL_{exp}. In (17-b), the nonthematic *it* is linked to SUBJ, *to me* to OBL_{exp}, and the subclause *that there is a problem with the proposal* to COMP. The empty subject argument can thus either be filled by a raised argument from an XCOMP, as in (17-a), or by the nonthematic *it*, as in (17-b).

While the typical analysis of the subject *it* in the *it*+subclause construction is as a non-thematic 'formal' subject, Berman (2003) goes against this idea in relation to German data. Berman gives two separate analyses for the subject *es* in the sentences in (18).

(18) a. weil es gesagt wurde, [dass Hans krank ist].
because it said was that Hans sick is
'because it was said that Hans is sick.'
b. weil es mich stört, [dass sie den Hans liebt].
because it bothers me that she Hans loves
'because it bothers me that she loves Hans.'

Berman (2003) argues that the sentence in (18-a) has a thematic subject *es* in conjunction with an adjunct subclause, i.e. *it*+ADJ, while the sentence in (18-b) has a non-thematic subject *es* in conjunction with a complement

subclause, i.e. *it*+COMP. The evidence given for this distinction is derived from *wh*-extraction. The analysis provides an explanation why it is not possible to extract out of the subclause in (19-a), while, according to Berman (2003: 152), extraction is acceptable from the subclause in (19-b).

(19) a. *Was wurde es gesagt, [dass er gelesen hat].
What was it said that he read has
'What was it said that he has read.'
b. Wen stört es dich, [dass sie liebt].
Who bothers it you that she loves
'Who does it bother you that she loves.'

In (19-a), passive *gesagt werden* takes a subject with the thematic role *theme*, corresponding to the pronoun *es*. The fact that the sentence in (19-a) is ungrammatical supports the analysis of the subclause as an adjunct. As is standardly assumed, adjuncts constitute syntactic islands out of which extraction is not possible, or is at least considerably more difficult than for complements (Bresnan et al., 2016: 287).

In (19-b), according to Berman (2003: 161), the verb *stören* takes a non-thematic subject *es*, an experiencer object and a *theme* complement. The *it*+COMP analysis accounts for the grammaticality of extraction out of the subclause, shown in (19-b).

As will be seen in Section 5.2.4., some aspects of Berman's analysis, such as the distinction between thematic and non-thematic subjects, are adopted in my analysis, while others are rejected. One problem with Berman's analysis is the fact that the presence of non-thematic *es* in Berman (2003) seems to a large extent to be the result of lexical idiosyncracies. How is it, for instance, that a verb such as *stören* requires a non-thematic subject? As will be seen, my analysis gives a different account of the German data, where the presence or absence of a non-referential subject is not taken to be the result of lexical idiosyncrasies.

4.3 Clausal subjects and the *it*+subclause construction in Old English

As has been shown, there are opposing views on the properties of clausal subjects and the *it*+subclause construction in Present-day English. The same is true with respect to studies on clausal subjects and the *it*+subclause construction in historical English, which is the topic of the present section.

4.3.1 The existence of clausal subjects in Old English

Let us start out with the question if subclauses could constitute subjects in early English. Among the studies discussing this issue we find Visser (1963-1973); Fischer & Van Der Leek (1983); Mitchell (1985a,b); Traugott (1992); Anderson (1997); Méndez Naya (1997); Zimmerman (2015). In the present section, I will focus on the last three of these studies, where Anderson (1997) argues that subclauses cannot constitute subjects in Old English, while Zimmerman (2015) and Méndez Naya (1997) both speak to the contrary.

Anderson (1997: 26) claims that 'there is no motivation for regarding any sentential arguments as subjects in Old English'. According to his view, this statement pertains to both structural and functional subjecthood, being valid for both *that*-clauses and infinitive phrases. The basis for this statement with respect to syntactic (structural) subjecthood is the claim that lexically governed subclauses do not occupy clause-initial position[7], and with respect to morpho-syntactic (functional) subjecthood that they do not participate in verb agreement or subject sharing. In the next chapter, these claims will be further discussed in relation to my data.

While Anderson (1997) claims that subclauses cannot constitute subjects in Old English, Méndez Naya (1997) argues that certain clause-final subordinate clauses do function as subjects during this period. Particularly, she claims that clauses depending on passive verbs taking accusative objects can be identified as subjects in Old English. Consider the sentences in (20) and (21).

(20) a. Petrvs se apostol awrat twegen pistolas,
Peter the apostle wrote two apostelic letters
'The apostle Peter wrote two apostelic letters.'
(colsigewZ,+ALet_4_[SigeweardZ]:928.384)

b. Twegen pistolas wurdon/sindon awriten
two apostelic letters were/are written
'Two apostolic letters were written.'
(Méndez Naya, 1997)

(21) a. Paulus awrat ðæt sio Godes lufu sie geðyld
Paul wrote that the God's love is patience
'Paul wrote that the love of God is patience.'
(Méndez Naya, 1997)

[7]Mitchell (1985b: 1) states for instance that 'OE noun clauses introduced by *þæt* never stand at the beginning of their sentence'.

b. Hit is awriten on Paules bocum ðæt sio Godes lufu sie
it is written in Paul's book that the God's love is
geðyld,
patience
'It is written in the book of Paul that the love of God is patience.'
(cocura,CP:33.215.21.1439)

In (20) and (21), two passive alternations with the verb *awritan* ('write') are given, one, in (20), in which the active sentence contains an accusative NP and one, in (21), which contains a complement *that*-clause. Méndez Naya (1997) seems to argue that the *that*-clause in (21-b) should be given the same analysis as the nominative subject in (20-b) on the grounds of the economy of the analysis. She writes that 'If the *þæt*-clause [...] is seen as equivalent to an accusative NP, I cannot think of any other plausible analysis for the embedded clause [...] than that of subject' (Méndez Naya, 1997). What immediately seems problematic about this analysis is that the presence of the nominative pronoun *hit* is not mentioned or discussed. Rather, a plausible analysis for this sentence is that the nominative *hit* constitues the subject and that the subclause constitutes an appositive adjunct. This is the analysis I propose in section 5.2.3 for similar sentences. I do think that the paralell between the behaviour of NPs and subclauses constitutes an argument in favour of the analysis of the subclause as subject, but only in those cases where there is no nominative pronoun *it*, such as in the sentence in (22).

(22) Be ðam is on bocum awriten þæt God þurh haligne
About those is in book written that God through holy
gast hine het faran to sumere mærre ceastre seo wæs
ghost them ordered travel to some great city that was
Niniue haten ...
Niniue called
'It is written in the book about them that God throught the holy ghost ordered them to go to some great city called Niniue.'
(coverhom,HomS_34_[ScraggVerc_19]:109.2485)

Unlike the sentence in (21-b), a construction such as the one in (22) seems to constitute a better passive counterpart to a sentence like the one in (21-a), supporting the idea that the subclause constitutes a subject.

As discussed in Zimmerman (2015), an alternative to the subclause occurring with an impersonal verb being analysed as a subject, is an

analysis where the subclause is analysed as an associate of an empty subject. Possible evidence that could be used to distinguish between these two analyses is the presence of a fixed subject position, and the possibility of extraction. If there is a fixed subject position, a clause-final subject would not be allowed and we would prefer the analysis where the subclause is an associate of an empty subject expletive. Likewise, if it is possible to extract from out of the subclause, the analysis of the subclause as a subject falls, as subjects typically constitute syntactic islands.

4.3.2 Non-nominative subjects in Old English

When discussing the status of clause-final subclauses with impersonal predicates in Old English, it is also important to say a few words about the phenomenon of non-nominative subjects. There is evidence to suggest that, when an impersonal predicate co-occurs with a dative experiencer, it is sometimes this dative experiencer that constitutes the subject. One proponent of this analysis is Allen (1986, 1995). Consider the sentence in (23), which is an example discussed in Allen (1986: 469).

(23) Feawum mannum gelimpð on þisum dagum, [þaet he
Few-DAT men-DAT happens in these days, that he
gesundfull lybbe hund-eahtatig geara]
healthy live eighty years
'It happens to few men in these days to live eighty years in health'.
(cocathom1,+ACHom_I,_32:458.216.6533)

In (23), there is a clause-initial dative phrase with an experiencer role, and a clause-final *that*-clause with a source role. Denison (1993: 64) says about sentences such as (23) that 'there is no obvious way, pretheoretically, of discriminating between types (i) and (ii)', i.e. to discriminate between the analysis where the dative experiencer is the subject, and where it is not. However, based on position, the complementary distribution between a dative phrase and a subject *hit*, and coordinate subject deletion, Allen (1995) argues that the dative phrase in sentences such as (23) can be established as subjects. In (24), an example is given of coordinate subject deletion, where a dative phrase seems to provide the subject of both predicates (Haugland, 2006: 330).

(24) Ða þeahhwæðere ofþuhte þam ælmihtigum gode
then however regretted that.DAT almighty.DAT god.DAT
ealles manncynnes yrmða & smeade hu he mihte ...
all mankind's miseries and pondered how he could ...

> 'Then, nevertheless, the Almighty God was grieved for the miseries of all mankind, and [he] meditated how he might ...'
> (cocathom1,+ACHom_I,_13:281.8.2355)

In (24), the dative phrase *þam ælmihtigum gode* ('the/that almighty god') is also the subject of the verb *smeagan* ('to ponder') in the second conjunct. Sentences such as (24) suggest that dative phrases with impersonal verbs can constitute subjects. In impersonal sentences with a dative phrase and a subclause, it also seems reasonable to sometimes assume that it is not the subclause but the dative experiencer that constitutes the subject. The subclause would then be analysed as a complement or an adjunct. The non-nominative subject hypothesis is further discussed in Chapter 5, Section 5.3.2, where this hypothesis is further supported.

4.3.3 The *it*+subclause construction in Old English

With respect to the data from Old English, there are, just as for Present-day English, different analyses as to the status of the subject *it* and the propositional subclause in the *it*+subclause construction. When it comes to the analysis of *it* as thematic or nonthematic, several studies treat it as non-thematic (Wahlén, 1925; Mitchell, 1985a; Anderson, 1988; Haugland, 2006). The same studies differ as to whether they treat the subclause as a complement or as an adjunct. Wahlén (1925) and Mitchell (1985a) analyse the subclause as being in apposition to the subject *it*, i.e. they provide an adjunct analysis. Anderson (1988) and Haugland (2006), on the other hand, go for the analysis of the subclause as a complement. In the present section, some of the the arguments for the two positions are presented.

With respect to Mitchell (1985), the analysis is problematic. Although he claims that the subclause in the *it*+subclause construction stands in apposition with a formal subject *it*, no arguments are given for this analysis (Mitchell 1985, §1039). Consider the sentence in (25).

(25) þa gelamp hit [þæt ðam gyftum win
then happened it that the wedding-gift-giving wine
ateorode]
was-lacking
'It then happened that there was no wine at the wedding'.
(cocathom1,+ACHom_I,_4:206.8.646)

Mitchell makes the assumption that the *that*-clause in (25) stands in apposition to a formal, i.e. nonthematic, subject *hit*, without discussing the matter further. As pointed out in Anderson (1988: 11), the analysis of the subclause in (25) as appositive seems to contradict Mitchell's (1985: §1428) own definition of apposition, where he adopts the following definition from Pei & Gaynor (1954): 'the use of paratactically joined linguistic forms occurring in the same clause or sentence which refer to the same referent and which have the same or similar grammatical form and function, but not the same meaning'. Assuming that a 'formal' subject does not have a referent, the analysis of the subclause in (25) as standing in apposition to the subject *hit* is not consistent.

Just like in Mitchell (1985a) and Wahlén (1925), the assumption that the subject *it* is a meaningless 'formal' subject is also supported by Haugland (2006: 400). One argument she gives in support of her analysis is that a subject *hit* in Old English is frequently not present in the relevant constructions. She writes that 'it is difficult to account for the low frequency of the overt pronoun in a language that did not normally omit referential subject pronouns'. Although this argument is reasonable[8], it does not preclude an LFG analysis of the null+subclause construction where the subclause constitutes the morphosyntactic subject. As will be seen in Section 5.3.2, I analyse the *it*+subclause construction in Old English as *it*+ADJ, i.e. with a thematic subject *it* and an adjunct subclause, when a subject *it* is present, and as only containing a clausal subject in those cases where there is no overt preverbal subject constituent. Support for my analysis is given there.

4.4 Summary

In this chapter, studies on the syntax of preposed clausal subjects and the *it+subclause* construction have been presented and discussed with

[8]In opposition to claims to the contrary (e.g. van Gelderen, 2000: 121), Rusten (2013: 990), based on an empirical survey, holds that '[v]ery low overall and text-individual frequencies for these subjects prompted the conclusion that previous accounts of the distribution, extent and "idiomaticity" of empty subjects in OE are unsubstantiated'.

respect to PDE and early English. With respect to the preposed clausal subject construction, tests were identified in the literature for the analysis of the syntactic and morphosyntactic properties of the subclause in this construction. In terms of structural position, there are different opinions as to the question if subclauses can occupy the structural subject position of English. In terms of morphosyntactic properties of the subclause in the preposed clausal subject construction, the data presented supports the analysis of the subclause as a subject.

With respect to the *it*+subclause construction in PDE further tests were identified for the analysis of the status of the subject *it* as thematic or non-thematic and the subclause as a complement or an adjunct. Based on examples given in the previous literature, there seems to be a possible difference between raising verbs and adjectival predicates in the status of the subject *it* and the subclause. Raising verbs seem to take a non-thematic subject *it* and a complement subclause, while non-raising predicates seem to take a thematic subject *it* and an adjunct subclause. In early English, the existence of subclauses as subjects is contested. Some studies do not recognise the existence of subclauses as subjects in early English, while others do. In the discussion, the existence of a fixed subject position, the possibility of extraction out of the subclause and the economy of the analysis have been seen as relevant.

The questions introduced in this chapter will be further discussed in the next chapter, where the data from the historical corpora is evaluated.

Chapter 5

Results and analysis

In Chapter 4, related studies on the syntax of clausal subjects and the *it*+subclause construction were discussed. In the present chapter, the questions introduced in Chapter 4 are tested against the material in the historical corpora. Section 5.1 concerns the analysis of preposed clausal subjects in Early and Late Modern English, Section 5.2 the *it*+subclause construction in Early and Late Modern English, and Section 5.3 both constructions in Old and Middle English.

5.1 Preposed clausal subjects in Early and Late Modern English

As previously discussed, there is a distinction to be made between functional and structural subjecthood, or, in the terms of Anderson (1997), between morphosyntactic and syntactic subjecthood. In the present section, the tests for functional and structural subjecthood, presented in Section 4.1 above, are applied to sentences containing a subclause annotated as a subject in the Early and Late Modern English corpora.

5.1.1 Functional subject properties

The functional subject properties from Section 4.1 are verb agreement, raising, control and coordinate subject deletion. Two of these properties, subject raising and coordinate subject deletion, show positive evidence for the subject status of subclauses, while the other two show inconclusive results in this respect.

As for subject raising, there are a number of examples from the corpora. The following sentences illustrate the subject raising construction in Late Modern English and Early Modern English, repectively.

(1) a. [To make a speech at the Literary Fund dinner] seems to be a duty expected from an ex-Prime Minister.
(TROLLOPE-1882,185.463)
b. And [that this may be so], seems with great probability to be argued from the strange Phaenomena of sensitive Plants, wherein [. . .]
(HOOKE-E3-H,116.100)

In (1-a) the subclause *to make a speech at the Literary Fund dinner* is the subject of both the main clause predicate *seem* and of *be a duty to be expected from an ex-Prime Minister*. In (1-b), it is the clause *that this may be so* that constitutes the subject of the passive construction *be argued*. There are also examples of clausal subject raising in conjunction with a passive predicate, a type of construction that is discussed in Section 5.2.2. Consider the sentences in (2).

(2) a. and [to acquire an acquaintance with their doctrines and systems] came to be considered as the most essential part of a liberal education.
(WHEWELL-1837,22.196)
b. [That a thing, so simple in itself, should abound with so many advantages,] is scarcely to be supposed, at a first glance;
(LANCASTER-1806,53.294)
c. [that the Public were so mad after his trash] is not to be wondered at,
(HAYDON-1808,1,33.846)

In the sentences in (2), the clausal subjects are also the subjects of a predicate in an infinitival clause. The fact that the subclauses in (1) and (2) participate in subject raising gives important support for the analysis of the subclause as a subject.

Also coordinate subject deletion shows positive evidence for the subject status of subclauses in Early and Late Modern English. Consider the sentence in (3), taken from the Late Modern English period.

(3) So different is the idiom of the Latin from that of the English, that [to translate with propriety from the one into the other, and especially from the latter into the former,] is thought a very difficult

task, and is believed by many to be more than can be expected from the tender years and the confined ideas of a school-boy. (CHAPMAN-1774,208.306)

In this sentence, the subclause *to translate with propriety from the one into the other, and especially from the latter into the former* forms the subject of both conjoined predicates. Since the subject of the first clause licenses the leaving out of the subject of the second clause, this phenomenon is commonly known as coordinate subject deletion.

While subject raising and coordinate subject deletion show positive evidence, control and verb agreement are inconclusive as tests in this context. With respect to control, no examples are found in the corpora of a clausal subject controlling the subject of an infinitival clause. Considering the fact that the preposed clausal subject construction as well as preposed clauses in general are relatively infrequent (see Tables 5.1 and 5.2, below), it is not altogether surprising that the combination of a preposed clausal subject and a control construction cannot be found in the corpus material.

With respect to verb agreement, consider the sentence in (4) from the Early Modern English period.

(4) [Whether from such Experiments one may argue, that *'t*is but, as *'t*were, by accident that Amber attracts another body, and not this the Amber; and whether these ought to make us question, if Electricks may with so much propriety, as has been hitherto generally supposed, be said to Attract], are doubts that my Design does not here oblige me to examine. (BOYLE-E3-H,20E.54)

This sentence contains two coordinated *whether*-clauses in clause-initial position. It is followed by the copula *be* in the plural and the plural NP *doubts that my Design does not here oblige me to examine*. If we assume that the clause-final NP is not the subject, it seems as if the two coordinated *whether*-clauses trigger plural marking on the verb. However, considering the presence of a plural NP, which could also be analysed as a subject, no conclusion can be reached about whether verb agreement between a clausal subject consisting of two coordinated subclauses and the predicate takes place in (4). In the corpus, there is no clausal subject consisting of two coordinated clauses occurring together with an adjectival predicate or a nominal predicate with an NP in the singular. When there is a verb in the plural in the preposed clausal subject construction, it in all attested instances in the corpora occurs in conjunction with a plural NP in the predicate.

Furthermore, there are also examples of coordinated subclauses tagged as subject, where the finite verb is in the singular. Consider the sentences in (5).

(5) a. [to love such a Person, and to contract such friendships,] is just so authorized by the principles of Christianity, as it is warranted to love wisdome and vertue, goodnesse and beneficence, and all the impresses of God upon the spirits of brave men.
(JETAYLORMEAS-E3-P2,47.231)

b. [To separate these studies, and to allow students to neglect one of them, because some persons have a taste for one, and some persons for another, of these lines of reading,] is to abdicate the functions of education altogether.
(WHEWELL-1837,41.401)

In (5-a), there are two coordinated infintival clauses in conjunction with an adjectival predicate, where the copula *be* is in the singular. In (5-b), there are also two coordinated infinitival clauses annotated as subjects, followed by the copula *be* in the singular and then another infinitival clause. Considering the fact that *be* in these examples is in the singular and not in the plural, it further weakens the claim that there is verb agreement between a clausal subject and the finite verb.

Summarising the above discussion, we may establish that subject raising and coordinate subject deletion support the analysis of different kinds of subclauses as subjects. However, there are no clear examples from control or verb agreement to be found. The lack of data from verb agreement is particularly problematic for the analysis of the subclause as a subject, considering the importance of verb agreement as a defining subject property.

5.1.2 Structural subject properties

In Present-day English, perhaps the most important property of subjects is their structural position. Subjects are typically associated with the Spec,IP position.

Considering the structural position of subordinate clauses generally, they typically follow (although not always immediately) the finite verb. Table 5.1 shows the number and proportion of clausal arguments preceding the finite verb in the historical corpora.
As can be seen from the table, there are relatively few clausal arguments preceding the finite verb, especially with respect to *that*-clauses. The

Table 5.1: Absolute frequencies and percentages for clausal arguments preceding the finite verb

		tokens	%
Old English	that	8	<1
	wh	18	<1
	inf	261	6
Middle English	that	65	1
	wh	76	4
	inf	145	2
Early Modern English	that	64	<1
	wh	221	10
	inf	404	3
Late Modern English	that	62	1
	wh	165	14
	inf	277	3

overwhelming majority of clausal arguments follow the finite verb in all historical periods. However, there also seems to be a change, which especially applies to *wh*-clauses. Interrogative clauses as arguments increasingly occur in a preverbal position, from less than one percent to fourteen percent between Old English and Late Modern English.

If we look at the frequency of clausal subjects throughout the history of English, further changes can be seen. In Table 5.2, we see the number of nonextraposed clauses for the different historical periods that are tagged as subjects in the corpora. Relative frequencies are also given in the form of number of occurrences per 100,000 clauses (clauses tagged as IP in the corpus).

Table 5.2 shows that the frequency of nonextraposed clausal subjects is low in all periods of the English language. There are, however, some developments to be noted. Infinitival clausal subjects show a gradual increase from below one instance per 100,000 clauses to 95 instances per 100,000 in the Late Modern English period. The other clause types, *that*-clauses and *wh*-clauses, also show a gradual increase, although not as dramatic. For infinitival clauses, the most prominent shift happens between the Old English and Middle English period; for *that*-clauses it happens between the Middle English and Early Modern English period; and, for *wh*-clauses, it happens between the Early Modern English and

Table 5.2: Absolute and relative frequencies for preposed clauses tagged as subjects in the historical corpora

		tokens	/100K clauses
Old English	that	1	<1
	wh	1	<1
	inf	2	<1
Middle English	that	7	3
	wh	6	3
	inf	45	21
Early Modern English	that	24	16
	wh	9	6
	inf	139	92
Late Modern English	that	48	38
	wh	27	21
	inf	121	95

Late Modern English period. Infinitival clauses are the most frequent clause type occurring as a nonextraposed subject throughout all periods of English. This is not altogether surprising, given the assumption that infinitives are the most noun-like of the types of subclauses, and that their historical origin is as a nominal category for the bare infinitive and as a prepositional phrase for the inflected infinitive with *to*[1](cf. Los, 2005).

Even though the clausal subjects in Table 5.2 are tagged as subjects in the corpus, it is not clear that they occupy the structural subject position at any stage of the language. As argued in Section 4.1., there are a number of environments indicative of structural subject status. Koster's (1978) positional evidence concerns the grammaticality of subclauses (i) in a subordinate clause, (ii) in subject-auxiliary inversion, (iii) after fronting, and (iv) in connection with so-called bisentential verbs.

Searching for these structures in the Penn Corpora of Historical English yields an interesting result. There is a difference between non-extraposed clausal subjects contained within subordinate clauses (i), where a number of examples are found, and the other environments (ii-iv), where almost no examples are found. No examples are found of subclauses after subject-

[1]Los (2005) discusses the categorial status of to-infinitives and concludes that they behave as non-finite subjunctive clauses rather than prepositional phrases in Old English (Los, 2005: Chapter 7).

verb inversion (ii) or in connection with bisentential verbs (iv). With respect to clausal subjects occurring after a lexically governed fronted constituent (iii), there is only one possible example to be found. This is the one given in (6), containing an infinitival clausal subject from the Early Modern English period.

(6) For, of all outward qualities, [to ride faire], is most cumelie for him selfe, most necessarie for his contrey,
(ASCH-E1-P1,10R.187)

Here, the PP *of all outward qualities* occurs in a fronted position followed immediately by the clausal subject.

While clausal subjects in subject-auxiliary inversion or following a fronted element are very rare, clausal subjects in subordinate clauses are more common. The following sentences illustrate their usage in the Late Modern period.

(7) There we discover, that [to sacrifice our intellectual, our moral enjoyments, to the lower and more inglorious propensities of our nature,] is, in reality, to inflict a heavy punishment on ourselves;
(CHAPMAN-1774,216.356)

(8) If the antiquary has a right to diversify his conjectures, the etymologist may fairly demand the same privilege; because, [to chain his fancy] would be to annihilate his existence, which is a barbarity that no individual can justify towards another.
(TURNER1-1799,52.261)

In (7), the clausal subject *to sacrifice our intellectual, our moral enjoyments, to the lower and more inglorious propensities of our nature* occurs as the first constituent within a *that*-clause; in (8), the clausal subject *to chain his fancy* occurs after the subordinator *because*.

As can be seen from these examples of subclauses in the typical subject position Spec,IP, they all contain infinitival clauses. Additionally, with respect to (7) and (8), the clausal subject is followed by the identificational *be* and another infinitival clause. Possibly, this observation can be connected to matters of weight and complexity, which are further discussed in Chapters 4 and 5. Many of the infinitival clausal subjects shown contain relatively few words.

In short, there is evidence suggesting that infinitival clauses can occupy the structural subject position in Early and Late Modern English. As for the other clause types, there is no support in the corpora for the hypothesis that these can occupy the structural subject position. Tentatively, the

three phrase structure rules in (9) are proposed for Early and Late Modern English.

(9) a. Preposing rule in EME and LME:

$$\begin{array}{lcccc} \text{CP} & \rightarrow & \text{XP} & & \text{C'} \\ & & (\uparrow \textsc{udf}) = \downarrow & & \\ & & (\uparrow \textsc{udf}) = (\uparrow \textsc{gf}) & & \end{array}$$

b. Subject rule in EME and LME:

$$\begin{array}{lcccc} \text{IP} & \rightarrow & \text{NP|IP} & & \text{I'} \\ & & (\uparrow \textsc{subj}) = \downarrow & & \end{array}$$

c. Alternative IP rule without subject:

$$\text{IP} \rightarrow \text{I'}$$

The rule in (9-a) says that any type of phrase, an XP, can occur in the fronted position in the specifier of the CP, bearing the grammatical function UDF (unbounded dependency function). Furthermore, this UDF is equated to any grammatical function, i.e. including SUBJ. This rule thus licenses preposed clausal subjects[2]. The rule in (9-b) says that an IP licenses two daughters, where the first daughter is an NP or an IP and the second daughter is an I'. The NP or IP furthermore provides the SUBJ of the clause. This rule thus licenses IP-subjects, for instance infinitival clause subjects, but not CP-subjects, such as *that*-clauses or *wh*-clauses, in the subject position of English, Spec,IP, in these periods. Lastly, the rule in (9-c) licenses an I' daughter of IP with no sister, which is required when a clausal subject occupies the Spec,CP position. Then, there will be no subject in the Spec,IP position.

5.2 The *it*+subclause construction in Early and Late Modern English

As discussed in Section 4.2, the preposed clausal subject construction alternates with a construction featuring a subject *it* in conjunction with a propositional subclause, i.e. the *it*+subclause construction. With respect to this latter construction, there is a discussion whether the subject *it* is thematic or non-thematic and whether the subclause constitutes a

[2]It should be noted that the rule in (9-a) also licenses preposed NP subjects. However, this possibility is left unexplored in this dissertation.

complement or an adjunct. In the present section, the properties of the *it*+subclause construction are discussed with respect to the material from the Early and Late Modern English corpora. The data presented in the present section are also discussed in Ramhöj (2015), where similar conclusions are drawn.

As a result of the investigation presented in this chapter, I propose that the *it*+subclause construction can be divided up into two different constructions: (i) the *it*+ADJ construction, with a thematic subject *it* and an adjunct subclause, and (ii) the *it*+COMP construction, with a non-thematic subject *it* and a complement subclause.

Recall from Section 4.2 that Shahar (2008: 38) found differences between predicates in the acceptability of extractions of an adjunct out of the subclause in the *it*+subclause construction. In particular, he reported a difference between raising predicates, *seem* and *be likely* and non-raising adjectival predicates, *be possible* and *be clear*. For convenience, the sentences are repeated in (10).

(10) a. ?How is it possible [that John kissed Mary]? (With passion)
b. ?How is it clear [that John kissed Mary]? (With passion)
c. How is it likely [that John kissed Mary]? (With passion)
d. How does it seem [that John kissed Mary]? (With passion)

These data suggest that, for raising predicates, such as *seem* and *be likely*, the subclause is a complement, while for non-raising adjectival predicates, such as *be possible* and *be clear*, the subclause is an adjunct in the *it*+subclause construction. Based on the fact that adjuncts are typically analysed as syntactic islands (Bresnan et al., 2016: 287), out of which extraction is not possible, the data suggest that the *it*+COMP construction is associated with raising predicates, while the *it*+ADJ construction is associated with non-raising predicates. How, then, does this distinction in the status of the subclause correspond to the distinction between a thematic and non-thematic subject *it*? The judgements presented by Kaltenböck (1999: 57), with respect to a non-raising adjectival predicate, suggested that the subject *it* for this predicate should be analysed as thematic, as it could be deleted under conjunction with a thematic subject *it*. If we apply the same test to a raising predicate, *seem*, the result appears to be ungrammatical. The sentences are repeated in (11).

(11) a. A: Have you heard of the bank robbery in London?
B: Yes, *it's* terrible, and (*it's*) hard [to believe that they actually got away with it].

b. A: Have you heard of the bank robbery in London?
B: Yes, *it's* terrible, but ?(*it*) seems [that they actually got away with it].

The judgements in (11) suggest that non-raising adjectival predicates have a thematic subject *it* in the *it*+subclause construction, while raising predicates have a non-thematic subject *it* in the *it*+subclause construction.

In conclusion, the judgements in (10) and (11) aim at different parts of the analysis of the *it*+subclause construction, but they both support the hypothesis that the *it*+COMP construction with a non-thematic subject *it* and a complement subclause is associated with raising predicates, while the *it*+ADJ construction with a thematic subject *it* and an adjunct subclause is associated with a non-raising predicate. This is also what is proposed in my analysis of the *it*+subclause construction.

In the following subsections, the present analysis is accounted for in more detail and discussed in relation to data from the Early and Late Modern English periods. Section 5.2.1 concerns the *it*+COMP construction, and Section 5.2.2, the *it*+ADJ construction. Lastly, Section 5.2.3 concerns a comparison with Present-day High German.

5.2.1 The *it*+COMP construction

In the analysis given here, the *it*+COMP construction in English occurs exclusively in connection with subject-to-subject raising verbs, including passive subject-to-subject raising constructions. In the present section, my analysis of raising verbs and the way in which they connect with the *it*+COMP construction is presented in more detail. My analysis of the argument structure of raising verbs is given in (12), relating both to lexical raising verbs, such as *seem* and *appear*, and to passive raising constructions.

(12) The argument structure of raising verbs:

	∅	proposition
	|	|
raising verb	⟨arg1	arg4⟩
	[–r]	[–o]
	|	|
	SUBJ	COMP/XCOMP

In (12), an abstract representation of the proposed argument structure for raising verbs is given. Raising verbs take a SUBJ which is not associated with a thematic role and a second argument which is linked to the role proposition and could alternatively be realised as a COMP or an XCOMP. Individual raising verbs might diverge slightly from this representation. The verbs *seem* and *appear*, for instance, take an optional experiencer argument, which is linked to arg4[–o], pushing the proposition role to arg5[–o]. The argument structure for *seem* would thus be represented as in (13).

(13) The argument structure of the verb *seem*:

	∅	experiencer	proposition
	|	|	|
seem	⟨arg1,	arg4,	arg5⟩
	[–r]	[–o]	[–o]
	|	|	|
	SUBJ	(OBL$_{exp}$)	COMP/XCOMP

The characteristic feature of raising verbs is that the arg1[– r] slot is not associated with a thematic role and that there is an XCOMP as one of the arguments. This is true both for lexical raising verbs, such as *seem*, and for passive raising constructions.

In Early and Late Modern English, the verb *seem* may enter into the following constructions, (i) the raising construction, (ii) the preposed clausal subject construction, and (iii) the *it*+subclause construction, all of which are captured by the argument structure in (13).

(14) a. He seems [to carry about with him the Fury of the Lion]. (BOETHPR-E3-P2,173.486)
b. and that first, because there seems [to be no other use of it]. (HOOKE-E3-P2,163.136)
c. And [to love God] seemed to him a presumptuous thing, ... (BURNETROC-E3-P1,53.114)

d. It seems to me [that the Athenian ideal - that of strong intellectual capacity - is left out of sight altogether]. (BENSON-1908,55.182)

In (14-a) and (14-b), we have two examples of the raising construction, where the subject of *seem* also constitutes the subject of the embedded predicate. In these two sentences, the subjects *he* and *there* correspond to the non-thematic arg1[–r] slot. As such, they don't have a thematic relation to *seem*, but instead to the embedded predicates within the infinitival clauses. A simplified f-structure for the sentence in (14-a) is given in Figure 5.1.

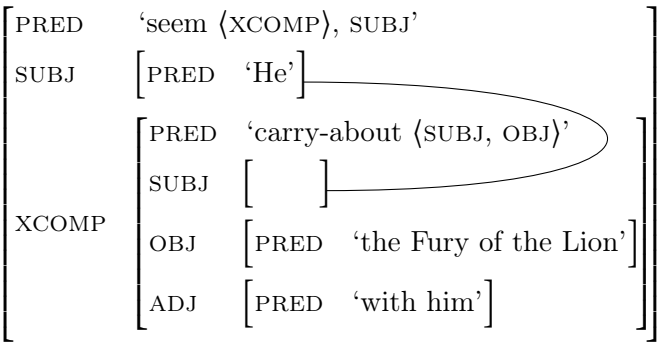

Figure 5.1: F-structure for the sentence *He seems to carry about with him the Fury of the Lion.*

As can be seen in the f-structure, there is a line between the subject of the main clause predicate *seem* and the subject of the embedded predicate *carry about*. This is a functional equation of the form (↑ SUBJ) = (↑ XCOMP SUBJ) which forms part of the lexical entry of the predicate *seem*. It says simply that the subjects of the two predicates are identical.

In (14-c), we have an example of a preposed clausal subject. Recall from Section 4.1. that preposed clausal subjects with the verb *seem* are only possible when there is a secondary predicate. Here, the secondary predicate is *(to be) a presumptuous thing*. Assuming that infinitival subject occurs in the structural subject position, as was suggested in Section 5.1.2, a simplified f-structure for the sentence in (14-c) is given in Figure 5.2.

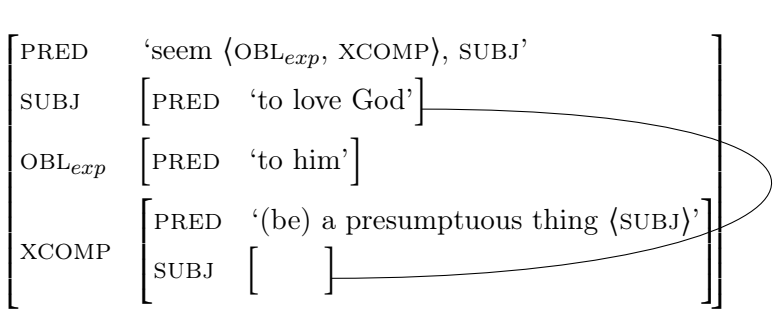

Figure 5.2: F-structure for the sentence *to love God seemed to him a presumptuous thing.*

Lastly, we have the *it*+subclause construction in (14-d). The *it*+subclause construction is here analysed as *it*+COMP, where the subject *it* is non-thematic and the subclause is a complement. A simplified f-structure for this sentence is given in 5.3.

$$\begin{bmatrix} \text{PRED} & \text{'seem } \langle \text{OBL}_{exp}, \text{ COMP} \rangle, \text{ SUBJ'} \\ \text{SUBJ} & \begin{bmatrix} \text{FORM} & \text{IT} \end{bmatrix} \\ \text{OBL}_{exp} & \begin{bmatrix} \text{PRED} & \text{'to me'} \end{bmatrix} \\ \text{COMP} & \begin{bmatrix} \text{PRED} & \text{'leave } \langle \text{SUBJ} \rangle \text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'the Athenian ideal'} \end{bmatrix} \\ \text{ADJ} & \begin{bmatrix} \text{PRED} & \text{'out of sight'} \end{bmatrix} \\ \text{ADJ} & \begin{bmatrix} \text{PRED} & \text{'altogether'} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Figure 5.3: F-structure for the sentence *It seems to me that the Athenian ideal [. . .] is left out of sight altogether..*

How do we know that the subclause in (14-d) has the grammatical function COMP? Evidence for this claim comes from *wh*-fonting and *wh*-extraction. The *wh*-fronting in (15-a) and the extraction in (15-b) represent the usage in the Late Modern English period.

(15) a. What a blessing did it seem [to have been permitted to accomplish a voyage, fraught with so many difficulties in the outset, but which, with the Almighty's blessing, we had in the sequel entirely escaped].
(MONTEFIORE-1836,128.5)
b. Q. In what state did it appear [to be at that time]?
(WATSON-1817,1,152.1839)

In (15-a), the complement of *seem* corresponds to the phrase *what a blessing*, which occurs in a fronted position. In (15-b), the phrase *in what state* corresponds to the complement of the copula *be* within the infinitival complement of *seem*. Since ADJs (adjuncts) typically are analysed as syntactic islands (Bresnan et al., 2016: 287), the occurrence of sentences such as those in (15) support the claim that a raising verb, such as *seem*, occurs in the *it*+COMP, where the subclause constitutes a COMP, and not in the *it*+ADJ construction.

Passive raising

Apart from lexical subject-to-subject raising verbs, such as *seem* and *appear*, another type of construction in which raising occurs concerns a group of passive predicates. Consider the sentences in (16).

(16) a. She is said [to have bine the death of her husband].
'She is said to have been the death of her husband.'
(MONTAGUE-E3-P2,1,219.78)
b. For the very reason why independence is sought is that it is judged good, and so power also, because it is believed [to be good].
(BOETHJA-1897,107.164)

In (16-a), the constituent *she* appears to be a thematic argument of the predicate *be the death of her husband*, rather than a thematic argument of the predicate *be said*. It is not *she* that is said, but rather the proposition that *she is the death of her husband*. Likewise, in (16-b), it is the proposition that 'power is good' that is believed and not the phrase *it*, referring to 'power'.

For the group of predicates occurring in the type of construction illustrated in (16), I assume the argument structure in (17). I will refer to this predicate as a passive raising predicate[3].

[3]The argument structure presented here differs from the one presented for the passive participle *said* in Ramhöj (2015).

(17) The argument structure of passive raising predicates:

	agent	∅	proposition
	|	|	|
passive raising predicate	⟨arg1,	arg2,	arg4⟩
	[–o]	[–r]	[–o]
	[+r]		
	|	|	|
	(OBL$_{agent}$)	SUBJ	XCOMP/COMP

Passive raising predicates, for instance *be said* or *be believed*, takes three arguments, arg1[–o], arg2[–r] and arg4[–o]. As with all passives, the arg1[–o] is assigned a [+r] feature, demoting arg1[–o] to the function of OBL$_{agent}$ (Kibort 2007). When arg1[–o, +r] is mapped to OBL$_{agent}$, arg2[–r] is mapped to SUBJ and arg4[–o] is mapped to XCOMP. For the purpose of illustration, a simplified f-structure associated with the sentence in (16-a) is given in Figure 5.4.

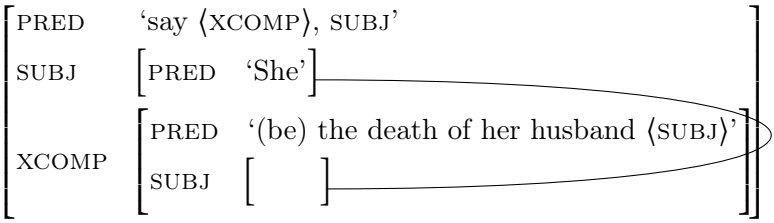

Figure 5.4: F-structure for the sentence *She is said to have bine the death of her husband*

This f-structure features the same functional equation as with a raising verb such as *seem*: the subject of the main clause predicate *be said* is equated with the subject of the embedded predicate *be the death of her husband*. A simplified version of the lexical entry of the passive raising predicate *be said* is given in (18).

(18) Lexical entry for the passive raising participle *said*:

said V (↑ PRED) = '*say* ⟨(OBL$_{agent}$), XCOMP⟩, SUBJ'
(↑ PASSIVE) = +
(↑ VFORM) = PAST PARTICIPLE
(↑ SUBJ) = (↑ XCOMP SUBJ)

The group of predicates that occur in the passive raising construction can be described as passivised propositional attitude verbs (Los, 2005). Postal (1974: 305-317) gives the following list of verbs participating in this kind of subject-to-subject raising for Present-day English:

(19) acknowledge, admit, affirm, allege, assume, believe, certify, concede, consider, declare, decree, deduce, demonstrate, determine, discern, disclose, establish, estimate, feel, figure, find, gather, grant, guarantee, guess, hold, imagine, intuit, judge, know, note, posit, presume, proclaim, prove, reckon, recognise, remember, report, reveal, rumour, say, show, specify, state, stipulate, suppose, surmise, take, think, understand, verify.

As can be seen, the verb *say* is one of the verbs in the list. One thing that should be noted about these verbs is that many of them, as for instance *say*, are reluctant to be realised in the correponding active argument structure of the one in (17), which is the subject-to-object raising structure (also known as ECM, exceptional case marking). While the verb *believe* occurs both in the active and passive sentences, the verb *say* here only seems to occur in the passive sentence.

(20) a. (i) Canst thou believe him [to be powerful], ...
(BOETHPR-E3-P2,113.163)
(ii) He is believed [to be powerful].
[constructed]
b. (i) *He said her [to be the death of her husband].
[constructed]
(ii) She is said [to have bine the death of her husband].
'She is said to have been the death of her husband.'
(MONTAGUE-E3-P2,1,219.78)

The verb *believe* in (20-a) occurs both in the active subject-to-object raising construction in (20-a-i) and the passive subject-to-subject raising construction in (20-a-ii). The verb *say*, on the other hand, only occurs in the passive subject-to-subject construction in (20-b-ii), and not in the active subject-to-object raising construction in (20-b-i). The argument structure of the active subject-to-object raising predicate is presented in (21).

(21) The argument structure of active subject-to-object raising verbs:

	agent	∅	proposition
	|	|	|
subject-to-object raising predicate	⟨arg1,	arg2,	arg4⟩
	[–o]	[–r]	[–o]
	|	|	|
	SUBJ	OBJ	XCOMP

As can be seen, active subject-to-object raising verbs are related to passive subject-to-subject raising verbs. However, certain verbs, such as *say* only occur in the passive alternant. The reason for this gap with respect to the verb *say* will be left to future research. Possibly, the semantics of the subject-to-object raising construction is not compatible with the lexical semantics of the active verb *say*.

Further support can be given for the claim that the passive subject-to-subject raising predicates really are raising predicates, and thus also occur in the *it*+COMP construction. Firstly, the subject of these predicates can constitute either of the non-thematic *it* or *there*, as exemplified in (22).

(22) a. It is said [that Dunkirk is sold to the French for four hundred thousand pound].
(HOXINDEN-1660-E3-H,280.184)

b. there is said [to be in it of Churches & Chappels, 150].
'There is said to be 150 churches and chapels in it (Prague)'
(JOTAYLOR-E2-P2,3,96.C2.299)

In (22-a), the subject of the main clause is the non-thematic *it*; in (22-b), it is the non-thematic *there*.

Secondly, this group of passive constructions seem to work in the same way as the rasing verbs *seem* and *appear* when it comes to extractions. Although no extractions in conjunction with the *it*+subclause construction were found in the historical corpora, googling the expression 'What is it' in conjunction with various past participles, numerous sentences from presumably trustworthy sources can be found for Present-day English. Three such sentences are given in (23).

(23) a. What is it said [that the eight planets represent]?[4]
b. What is it assumed [that humanity would do with such a key].[5]
c. From what is it claimed [that Victoria crosses are made]?[6]

In these sentences, the initial *wh*-phrases are extracted from within the *that*-clauses. The possibility of extraction here can be given as support for the hypothesis that passive constructions in the *it*+subclause construction are to be analysed as *it*+COMP.

5.2.2 The *it*+ADJ *construction*

As mentioned above, the *it*+COMP construction in in Early and Late Modern English is only used in connection with raising verbs, including the passive raising construction. With respect to non-raising predicates, a subject *it* is in this dissertation always analysed as thematic and the subclause as an adjunct, i.e. the *it*+ADJ construction. In this subsection, the analysis of the *it*+ADJ construction is illustrated with the verb *appear*, which can be analysed as either a raising verb or a non-raising intransitive verb.

The verb *appear* has more than one interpretation; hence, there are also two analyses, two lexical entries for this verb depending on meaning. The two options available are exemplified in (24).

(24) a. *appear*$_1$ ('to show itself')
And a vision appeared to Paul in the night:
'And a vision showed itself to Paul in the night.'
(AUTHNEW-E2-P2,16,1A.1072)
b. *appear*$_2$ ('to seem')
The children appeared [to be struck with amazement],
'It seemed that the children were struck with amazement.'
(COOK-1776,29.535)

The sentence in (24-a) illustrates the use of *appear*$_1$, which has the approximate interpretation 'to show itself', acting syntactically as an intransitive verb. The sentence in (24-b) illustrates the use of *appear*$_2$ with the approximate interpretation 'to seem, to give the impression of being'. With this second usage, the verb *appear* acts as a raising verb. The corresponding argument structures are given in (25).

[4] http://www.ultima-universe.com/u5walkthrough23.htm
[5] https://www.laetusinpraesens.org/musings/numb37.php
[6] http://www.militaryhistorytours.co.uk/category/uncategorized/

(25) a. The argument structure of *appear*$_1$ ('to show itself'):

	theme
	|
appear$_1$ ('to show itself')	⟨arg1⟩
	[–r]
	|
	SUBJ

b. The argument structure of *appear*$_2$ ('to seem'):

	∅	proposition
	|	|
appear$_2$ ('to seem')	⟨arg1	arg4⟩
	[–r]	[–o]
	|	|
	SUBJ	XCOMP/COMP

The different argument structures of *appear*, as given in (25), explain the differences in grammaticality between the sentences in (26) and (27) below.

(26) *appear*$_1$ ('to show itself')

a. [that in this matter I was not led by hym], very well and plainly appereth,
(MROPER-E1-P1,521.98)

b. - it plainly appeared by this time [that he had got a stiff neck, as he never once more turned].
(COLLIER-1835,13.370)

(27) *appear*$_2$ ('to seem')

a. - So it appears [to be].
(BOETHRI-1785,119.197)

b. *[To be so] appears.
[constructed]

The sentences in (26) and (27-a) are attested in the Late and Early Modern English corpora (Kroch et al., 2005, 2010), respectively, while the sentence in (27-b) is constructed. The ungrammaticality of sentences such as the one in (27-b) is frequently reported with respect to Present-Day English (e.g. Seppänen & Herriman, 2002), a judgement further supported by the fact that such sentences are not found in the corpora. In (26), we see that the verb *appear*$_1$ ('to show itself') occurs in either the preposed clausal

subject construction or the *it*+subclause construction. In (27), the verb *appear*$_2$ ('to seem') is ungrammatical[7] in the preposed clausal subject construction, as illustrated in (27-b).

The fact that a propositional subclause embedded under a raising verb, including the verbs *appear*$_2$ ('to seem') and *seem*, is ungrammatical without a secondary predicate has been referred to as obligatory extraposition (for further discussion, see Seppänen & Herriman, 2002). One relatively recent account of obligatory extraposition, Alrenga (2005), mentioned in Section 4.1, connects the ungrammaticality of the structure in (27-b) to the complement selection of the verb. In particular, Alrenga (2005: 196) argues that a verb such as *seem* 'only subcategorises for a CP complement (seem: [_ CP])'. On my account, which differs from Alrenga (2005), the ungrammaticality follows from the principle of completeness (Bresnan et al., 2016), given the argument structure assumed in (25-b). In (27-b), the verb *appear* ('to seem') does thus not have all the arguments it selects for. Given this analysis, no idiosyncratic selection for the syntactic category of the complement is required.

In (26), the interpretation of *appear* is 'to show itself', i.e. *appear*$_1$. With this interpretation, *appear* only selects for one argument, arg1[–r], which is mapped to SUBJ. This accounts for the fact that the sentence in (26-a) is grammatical, despite the fact that it superficially looks similar to the sentence in (27-b). In (26-b), where the subclause co-occurs with a subject *it*, we have the *it*+ADJ construction, which is the analysis given for all non-raising predicates, where a subject *it* occurs in conjunction with a propositional subclause. A simplified f-structure for the sentence in (26-b) is given in Figure 5.5.

[7] As pointed out for instance in Seppänen & Herriman (2002), the verbs *seem* and *appear* are grammatical in the preposed clausal subject construction if there is a secondary predicate, as exemplified in sentences such as *[That we were to take care of the remaining work] appeared (seemed) like a good idea at the time.*

$$\begin{bmatrix} \text{PRED} & \text{'appear } \langle \text{SUBJ} \rangle \text{'} \\ \text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'it'} \\ \text{ADJ} & \left[\text{PRED} \quad \text{'that he had got a stiff neck'} \right] \end{bmatrix} \\ \text{ADJ} & \left[\text{PRED} \quad \text{'plainly'} \right] \\ \text{ADJ} & \left[\text{PRED} \quad \text{'by this time'} \right] \end{bmatrix}$$

Figure 5.5: F-structure for the sentence *it plainly appeared by this time [that he had got a stiff neck].*

The figure shows that *appear*$_1$ takes one argument, the SUBJ. The subclause functions as an ADJ (adjunct) to the thematic subject.

A consequence of the analysis of a sentence as *it*+ADJ is that extraction out of the subclause should not be possible, or at least be considerably worse than for extraction out of a complement. We would thus predict that the sentence in (28) with *appear*$_1$, a variant of (26-a), is considerably worse than the sentence in (29-b) with *appear*$_2$. This also seems to be the case.

(28) *By whom does it appear well and plainly [that I was not led in this matter]?
[constructed]

(29) a. It appears [that Cobham took Raleigh to be either a God, or an Idol].
(RALEIGH-E2-P1,1,213.46)
b. What does it appear [that Cobham took Raleigh to be]?
[constructed]

The extraction in (28), where we have *appear*$_1$ ('to show itself'), seems to be considerably worse than the extraction in (29-b), where we have *appear*$_2$ ('to seem'). The fact that such extraction is considerably worse in comparison with raising verbs is also supported by the extraction data provided by Shahar (2008), shown above in the present section.

5.2.3 Comparison with Present-day High German

The two previous sections have given an account of the way the *it*+subclause construction can be divided into two types in Early and Late

Modern English: (i) the *it*+COMP construction and the *it*+ADJ construction. In Section 4.3, we saw that Berman (2003) also makes a distinction between *it*+COMP and *it*+ADJ with respect to various predicates in German, although she does not call the constructions by these names. In this subsection, an analysis is given of the data provided by Berman (2003). The analysis differs from that of Berman, and treats all instances of the *it*+subclause construction in German as *it*+ADJ, i.e. a thematic subject *es* in conjunction with an adjunct subclause. The sentences in (30) and (31), show the judgements for the presence or absence of *es* in conjunction with a propositional subclause. The sentences in (30-a) and (31-a) show the *it*+subclause and the null+subclause constructions embedded in a subclause, the sentences in (30-b) and (31-b) show *wh*-extractions out of the subclause, and the sentences in (30-c) and (31-c) show the preposed clausal subject construction for these predicates.

(30) Judgements for the passive *gesagt werden*

a. weil (es) gesagt wurde, [dass Hans krank ist].
because it said was that Hans sick is
'because it was said that Hans is sick.'

b. Was wurde (*es) gesagt, [dass er gelesen hat].
What was it said that he read has
'What was it said that he has read.'

c. [Dass Hans krank ist], wurde (*es) gesagt.
that Hans sick is was it said.
'That Hans is sick, was said.'

(31) Judgements for the verb *stören*

a. weil (es) mich stört, [dass sie den Hans liebt].
because it bothers me that she Hans loves
'because it bothers me that she loves Hans.'

b. Wen stört *(es) dich, [dass sie liebt].
Who bothers it you that she loves
'Who does it bother you that she loves.'

c. [Dass sie den Hans liebt], stört (*es) mich.
that she Hans loves bothers it me.
'That she loves Hans, bothers me.'

According to Berman (2003), a subject *es* is optional for both predicates. However, as can be seen in (30-b) and (31-b), the two predicates differ regarding their tendency to allow extraction out of the subclause. According to Berman, the passive *gesagt werden* allows extraction only when *es* is not present, while *stören* allows extraction only when *es* is present. In

contrast, the sentences in (30-c) and (31-c) show that, in the preposed clausal subject construction, it is ungrammatical for both predicates to insert a subject *es*.

With respect to the judgements in (30) and (31), I propose the argument structures in (32) for passive *gesagt werden* and the verb *stören*.

(32) a. The argument structure of passive *gesagt werden*:

	agent	proposition
	⟨arg1	arg4⟩
gesagt werden	[–o]	[–o]
	[+r]	
	(OBL$_{agent}$)	SUBJ

b. The argument structure of passive *gesagt werden* with demotion of the propositional argument:

	agent	proposition
	⟨arg1	arg4⟩
gesagt werden	[–o]	[–o]
	[+r]	[+r]
	(OBL$_{agent}$)	COMP

c. The argument structure of the verb *stören*:

	cause	experiencer
	⟨arg1	arg3⟩
stören	[–r]	[+o]
	SUBJ	OBJ$_{exp}$

First, it should be noted that the passive construction for German is analysed differently than the passive in Early and Late Modern English, discussed in Section 5.2.1. The claim that the analysis of the passive construction in German is different from the analysis of English is based on the fact that the relationship between the passive construction and the raising construction is different in German. In German, it does not seem as if the passive raising construction, exemplified for English in (16-a), is grammatical. Consider the sentence in (33).

(33) *Sie wird gesagt, [der Tod ihres Mannes gewesen zu sein].
She is said the death her man been to be
'She is said to have been the death of her husband' [constructed]

I take the ungrammaticality of (33) as an indication that the passive constructions in German and English are different. Accordingly, I argue that *werden* is not a raising verb in German, as represented in the argument structures in (32-a) and (32-b). A consequence of this analysis is that only the *it*+ADJ construction is available for passive constructions in German, as the *it*+COMP construction is dependent on the the presence of a raising verb.

In the case of (30-a), I assume that when *es* is present, this corresponds to the argument structure in (32-a), where the proposition is mapped to SUBJ, which is realised as a thematic subject *es*. The subclause is an ADJ, associated with the subject *es*. When *es* is not present, I assume the argument structure in (32-b), where the propositional argument, now corresponding to the subclause, is demoted to COMP.

Following Berman (2003: 156), I take it that the ungrammaticality of extraction in (30-b), where a subject *es* is present, is a result of the fact that the *dass*-clause is an adjunct, a syntactic island. The possibility of extraction when no *es* is present follows from assuming the argument structure in (32-b), where the propositional subclause is demoted to COMP.

For the preposed clausal subject construction in (30-c), I assume that demotion of the propositional argument is not possible, and that the propositional argument is mapped to SUBJ. Since there already is a subject, adding a subject *es* makes the sentence ungrammatical, thus violating the principle of uniqueness.

With respect to the verb *stören*, the possibility of extraction in (31-b) is unexpected, given the fact that *stören* is not a raising verb in German. With my argument structure analysis in (32-c), extraction would be expected to lead to ungrammaticality both with and without a subject *es*. With a subject *es* the subclause is an *adjunct*, and without a subject *es*, the subclause constitutes the subject. Both subjects and adjuncts are syntactic islands, which should make extraction ungrammatical in both cases.

Even though adjuncts are typically islands, as pointed out by Shahar (2008: 37), extraction out of adjuncts is sometimes possible. For Shahar (2008: 38), extraction of an adjunct out of a subclause constitutes a better test than extraction of a complement out of the subclause. Consider the constructed sentences in (34).

(34) a. Es stört mich [dass sie so laut spricht].
it bothers me that she so loud speaks
'It bothers me that she speaks so loud.'
b. ?Wie stört es dich, [dass sie spricht]?
how bothers it you that she speaks
'How does it bother you that she speaks.'

Consulted speakers of German have a hard time to interpret the adjunct *wie* in (34-b) as being extracted and associated with the predicate *sprechen* ('to speak') in the subclause rather than being associated with the predicate *stören* ('to bother') in the main superordinate clause.

Further evidence for the analysis of the subclause in (31-b) as an adjunct when *es* is present comes from the way in which other types of extraction structures in connection with the verb *stören* are interpreted when they occur in naturally occurring discourse. Consider the sentence in (35), taken from Alejchem (1922).

(35) Was stört es dich, [dass das Kind spielt]?
what bothers it you that the child plays
'What does it bother you that the child plays?'

The sentence in (35) has two interpretations. On the first interpretation, *was* functions as the complement of *spielen* in the subclause, while, on the second interpretation, it functions as the complement of the verb *stören* in the main clause, in an interpreation that could be described as exclamative. The actual interpretation in the text is the second one, where *was* is not the complement of *spielen*. In a search for the string 'Wen stört es' in a subcorpus of the *Deutsches Referenzkorpus*[8], all (27 out of 27) instances represent this second interpretation. If sentences such as (35) are consistently interpreted as non-extractions, giving support to the hypothesis that the subclause must be considered to be an adjunct.

Lastly, in the preposed clausal subject construction in (31-c), *dass sie den Hans liebt, stört (*es) mich*, the ungrammaticality caused by the presence of *es* supports the analysis where the subclause is analysed as subject, just like in the analysis of (30-c).

The German null+subclause construction in (30-a), where there is a clause-final subclause without a subject *es*, provides an interesting parallell to similar constructions in English. As can be seen, when there is a clause-final subclause in Present-day German no structural, i.e. syntactic, subject is required. In English, the null+subclause construction was common

[8]http://www1.ids-mannheim.de/kl/projekte/korpora/

before the widespread loss of nominal and verbal morphology during the Middle English period. However, there are occasional instances surviving in Early and Late Modern English. By way of illustration, consider the sentences in (36). The sentence in (36-a) derives from the Late Modern English corpus, while the sentences in (36-b) and (36-c) derive from the Early Modern English corpus.

(36) a. To these things must be added, [that moral Obligations can extend no further than to natural Possibilities]. (BUTLER-1726,241.108)
b. The viij day of Feybruarij was commondyd by the quene and the bysshope of London [that Powlles and evere parryche that thay shuld syng Te Deum Laudamus], (MACHYN-E1-P1,55.104)
c. and heere is to bee noted [that the first word a Nurse or a Mother doth teach her children if they bee Males, is Drinke, or Beere] ... (JOTAYLOR-E2-P1,3,78.C2.31)

In the sentences in (36), there is a clause-final subclause, but not a subject *it*. The fact that the sentences in (36) lack a structural subject is unexpected, given the importance of structural position as an indication of subjecthood during these periods. What accounts for the grammaticality of the sentences in (36) in the Early and Late Modern English periods will have to be a matter of future research.

5.3 Clausal subjects and the *it*+subclause construction in Old and Middle English

In the present section, the status of preposed clausal subjects and the *it*+subclause construction in Old and Middle English is discussed in relation to the corpus data. The first subsection concerns the existence of preposed clausal subjects in Old and Middle English, and the second subsection concerns the *it*+subclause construction in Old and Middle English.

5.3.1 Preposed clausal subjects in Old and Middle English

In the Old English corpus, there are four sentences containing subclauses annotated as subjects. Recall that clause-final subclauses are never

annotated as subjects in the historical corpora. The four preposed clausal subjects in the Old English corpus include one clause-medial *that*-clause, one clause-initial *wh*-clause and two clause-initial infinitival clauses. These four sentences are given in (37).

(37) a. & eft is [ðæt mon blissige & ne blissige] ðæt mon
and again is that man rejoice and not rejoice that man
ahebbe his mod of ðissum eorðlican to ðæm hefonlican,
arise his mind of this earthly to the heavenly
'and once again, the fact that a man rejoices and yet does not rejoice means that he exalts his mind from this earth towards the heavens.'
(cocura,CP:51.395.23.2685)

b. [Hwylc hire mægen wære], ma æfter hire deaðe
Which her power be, more after her death.DAT
gecyðed wæs.
revealed was.
'The extent of her virtue became more conspicuous after her death.'
(cobede,Bede_3:6.176.2.1718)

c. [To sittanne on mine swiþran healfe oððe on wynstran]
to sit on my right side or on left
nys me inc to syllanne ac þam þe hyt
NEG-is me.DAT you.ACC to give but them that it
fram minum Fæder gegearwod ys.
from my father prepared is.
'To sit on my left or right side is not for me to grant. Instead it is given to those for whom it has been prepared by my Father'
(cowsgosp,Mt_[WSCp]:20.23.1355)

d. [Fulwian þonne þæt cennende wiif oðþe þæt bearn
To baptise when that pregnant woman or that child
þæt þær acenned bið, gif heo syn þreade mid
that there begot is, if she is threatened with
frecernisse deaðes, ge heo in þa seolfan tiid þe heo
danger death's, either she in the very time that she
cenneð ge þæt þær acenned bið], nænige gemete
gives-birth or that there begot is, none at all manner
is bewered.
is prohibited.

'To babtise a woman who is pregnant or a new-born child, if either the woman at the very time she gives birth or the child is threatened with danger of death, is in no way prohibited.' (cobede,Bede_1:16.76.19.709)

As these four sentences are the only examples found of constituents tagged as clausal subjects in the Old English corpus, a closer inspection of them is required.

With respect to the two sentences in (37-b) and (37-d) from the translation of Bede's *Ecclesiastical History* as well as for the translation of the bible passage in (37-c), the preposed clausal subjects here should probably be regarded as loan syntax, rather than as 'natural' Old English. Consider first the two Latin correpondences in (38) from the *Ecclesiastical History* of the sentences in (37-b) and (37-d). The Latin sentences are taken from Colgrave & Mynors (1969: 240).

(38) a. [quae cuius esset virtutis], magis post
and-this what-kind-of be.CONJ virtue.DAT more after
mortem claruit.
death.DAT became-clear
'And of what virtue she was, became more clear after (her) death'

b. [Baptizare autem uel enixam
to-babtise on-the-other-hand either giving-birth
mulierem uel hoc quod genuerit, si mortis
woman.ACC or that what is-given-birth-to if death.GEN
periculo arguetur, uel ipsam hora eadem qua
danger.ABL proves or the-very-one time same that
gignit, uel hoc quod gignitur eadem qua natum est],
gives-birth or that what be-born same that born is
nullo modo prohibetur.
no way is-forbidden
'To babtise a woman who is pregnant or a new-born child, if either the woman at the very time she gives birth or the child is threatened with danger of death, is in no way prohibited.'

If we compare the Latin orginals in (38) with the Old English translations in (37-b) and (37-d), it turns out that the word order is more or less that of the Latin originals. In (38-a), apart from the relative order of the copula *be* (*esset* and *wære*) and the noun *virtue* (*virtutis* and *mægen*), the word order is exactly the same. This is even more apparent in the

long and complex sentence of (38-b), where the order of the constituents is more or less the same in Latin and Old English.

Consider also the Latin sentence in (39), which corresponds to the Old English sentence in (37-c). The Latin sentence is taken from the Vulgate Bible[9].

(39) [sedere autem ad dexteram meam vel sinistram] non est
to-sit however to the-right my or the-left no is
meum dare vobis, sed quibus paratum est a Patre
me.ACC given you.DAT but those-who prepared is by father
meo.
mine
'To sit on my left or right side is not for me to grant. Instead it is given to those for whom it has been prepared by my Father.'

Just like for the sentences from Bede's *Ecclesiastical History*, the Old English translation of the sentence in (39) is very close to the original. All phrases are in the same order, except for the relative order of the infintival *to give* (*to sylanne* and *dare*) and the second person plural pronoun *you* (*inc* and *vos*).

The sentences in (37-b)-(37-d) are translations that have a word order that more or less copies that of the Latin. The sentence in (37-a) also constitutes a translation from a Latin text that has a considerable proportion of word-by-word translations (Brown, 1969: 666). All considered, it is reasonable to assume that these sentences cannot be taken as support for the grammaticality of the preposed clausal subject construction in 'natural' Old English, but should probably be seen as Latin calques.

Thus, we have seen that there are no credible examples of preposed clausal subjects to be found in the Old English period. When we reach the Middle English period, a different picture emerges. In the Middle English corpus, we find 59 instances of the preposed clausal subject construction. Most of these are preposed infinitival clauses (45/59). Several of the earliest attested instances of the preposed clausal subject construction in Middle English come from the text *Ancrene Riwle*, which is estimated to have been composed between 1215 and 1222. It features the first preposed infinitival clause subject and the first preposed interrogative clause subject.

[9]http://www.sacred-texts.com/bib/vul/index.htm

(40) a. [Naut ane monglin honden. ach putten honden
Not one intermingle hands but put hands
utward bute hit beo for neode]; is wowunge efter godes
outward but it be for need is wooing for God's
grome & tollunge of his eorre.
fury and courting of his ire
'Not only to intertwine hands, but to put the hand outward, unless it be for necessity, is to court God's fury and to attract his anger.'
(CMANCRIW-1,II.92.1110)

b. [Hu god is to beon ane] is bað e i þe alde laʒe & i
How God is to be one is both in the old law and in
þe neowe isutelet.
the new revealed
'How God is to be one is revealed in both the Old and the New Testament.'
(CMANCRIW-1,II.121.1543)

The sentence in (40-a) contains a preposed infinitival clause subject and the sentence in (40-b) a preposed *how*-clause subject.

The first examples of a preposed *that*-clause subject that do not occur in texts that are word-by-word translations from other languages occur in the English Wycliffite Sermons from about the year 1400, which is considerably later than for *wh*-clauses or infinitival clauses. Two examples are given in (41).

(41) a. [Þat þes seruauntis telde þis kyng þat in þe seuenþe howr feuer forsooke þis child] bytokneþ a greet witt, as Robard of Lyncolne scheweþ.
'That the servants told this king that, in the seventh hour, fire forsook this child signifies great wisdom, as Robard of Lincoln shows'
(CMWYCSER,307.1439)

b. and [þat Crist towchede þis leprous] techeþ vs now þat þe manhede of Crist was instrument to his godhede, for to do myracles þat he wolde weren done;
'And that Christ touched this leprous man teaches us now that Christ's being a man was an instrument of his goodness for him to do the miracles that he wanted to be done.'
(CMWYCSER,364.2464)

The examples here show preposed *that*-clause subjects in connection with the transitive verbs *bitoknen* ('to signify') and *techen* ('to teach').

In the present section, the attested examples of preposed clausal subjects in Old and Middle English have been discussed. It seems as if the construction starts to occur in non-translations in Middle English; infinitival clauses and interrogative clauses can be found in early Middle English while *that*-clauses are first attested in texts from around the year 1400. In the next section, we turn to the analysis of the *it*+subclause construction in Old and Middle English.

5.3.2 The *it*+subclause construction in Old and Middle English

The present section[10] concerns the properties of the *it*+subclause constructions in Old and Middle English. The discussion will be limited to the two Old English verbs *þyncan* ('to seem') and *gelimpan* ('to happen') as well as their semantic equivalents *seem* and *happen* in Middle English. The reason for the choice of *þyncan* and *gelimpan* is that they are the most frequent verbs in conjunction with a propositional subclause in the Old English corpus. There are 321 instances of the verb *gelimpan* with a propositional subclause, and 112 instances of the verb *þyncan*. In total, there are 705 instances of the *it*+subclause construction and 1704 instances of the null+subclause construction in the Old English corpus. While *þyncan* and *gelimpan* are the semantic equivalents of *seem* and *appear*, they do not seem to have the same argument structure. The verbs *seem* and *appear* that, when they participate in the *it*+subclause construction take a nonthematic subject and a complement subclause. The verbs *þyncan* and *gelimpan* on the other hand, as will be demonstrated in the present section, do not participate in the raising construction, and, when they participate in the *it*+subclause construction, this construction is to be analysed as *it*+ADJ, and not *it*+COMP.

The structure of the discussion is as follows. First, the proposed argument structures for the verbs are presented. Support for the analyses is given in the form of extraction data, the (non-)existence of raising, and the co-occurrence patterns between a subject *it* and a dative experiencer.

The proposed argument structures for the verbs *þyncan* and *gelimpan* are presented in (42) and (43), respectively. Two argument structures are given for each verb. The choice between argument structures is assumed to follow from the mapping between thematic roles and argument slots.

[10]The data presented in the present section are also discussed in Ramhöj (2015), where similar conclusions are drawn.

An important part of the analysis of these verbs concerns the notion of dative subjects. As is discussed in Allen (1995), see Section 4.3.2, there is evidence, for instance from conjunction reduction, supporting the analysis of certain dative phrases in conjunction with intransitive verbs as subjects. More support for this analysis is given below.

The argument structures for the verb *þyncan* are given in (42).

(42) Argument structures of *þyncan*:

a. With dative subject (NP_{dat}+*verb*+[*that* . . .]):

	experiencer	proposition
	|	|
þyncan$_1$	⟨arg1	arg4⟩
	[–o]	[–o]
	|	|
	SUBJ	COMP

b. With dative object (*it*+*verb*+NP_{dat}+[*that* . . .]):

	proposition	experiencer
	|	|
þyncan$_2$	⟨arg1	arg3⟩
	[–r]	[+o]
	|	|
	SUBJ	OBJ

In (42-a), *þyncan*$_1$ takes two arguments, arg1[–o] and arg4[–o], which are mapped to SUBJ and COMP, respectively. In (42-b), *þyncan*$_2$ also takes two arguments, but now they are arg1[–r] and arg3[+o], mapped to SUBJ and OBJ, respectively. The differences depend on what argument slot (arg_n) the thematic roles experiencer and proposition are mapped to.

The argument structures for the verb *gelimpan* are given in (43).

(43) Argument structures of *gelimpan*:

a. With dative subject (*NP_{dat}+verb+⌈that . . .⌋*):

	experiencer	proposition
	|	|
gelimpan$_1$	⟨arg1	arg4⟩
	[–o]	[–o]
	|	|
	SUBJ	COMP

b. Without experiencer argument (*(it)+verb+⌈that . . .⌋*):

	proposition
	|
gelimpan$_2$	⟨arg1⟩
	[–r]
	|
	SUBJ

The argument structure for *gelimpan*$_2$ in (43-b) is identical to that shown for *þyncan*$_1$ in (42-a). In (43-b), the verb *gelimpan*$_2$ takes just one argument, arg1[–r], which is mapped to SUBJ. In the second argument structure given, there is thus no experiencer among the thematic roles selected.

Let us proceed to the data supporting the adoption of the argument structures in (42) and (43). First, let us consider the co-occurrence of a subject *it* and a dative experiencer for *þyncan* and *gelimpan*, when they occur with a propositional subclause. The frequencies are shown in Table 5.3, given per 100,000 clauses[11], with the token frequency given within parentheses.

Table 5.3: Relative (per 100K clauses) and absolute frequencies for the co-occurrence of *it* and NP-DAT.

	both *it* and NP-DAT	*it*	NP-DAT	neither *it* nor NP-DAT
þyncan	<1 (2)	0 (0)	46 (109)	<1 (1)
gelimpan	0 (0)	97 (228)	13 (30)	27 (63)

[11]The YCOE contains 236,046 IPs, where each IP represents a clause.

Table 5.3 shows that the verb *þyncan* more or less consistently occurs together with a dative experiencer argument and a subclause. There are very few cases in which *þyncan* takes with a subject *it*. This supports the argument structure in (42-a), where *þyncan* takes two arguments, a dative experiencer subject and a propositional complement in the form of a subclause. However, there are also two cases in which *þyncan* occurs with both a dative experiencer and a subject *it*. One of these is given in (44).

(44) Wel geradlic hyt eac þingð us [þæt we herto
well appropriate it also seems us.DAT that we hereto
gecnytton þa epactas],
tied those epacts
'It seems very appripriate to us that we tied the epacts to this'.
(cobyrhtf,ByrM_1_[Baker-Lapidge]:1.2.291.403)

In (44), *þyncan* takes both a subject *it* and a dative experiencer, which is here analysed as an object. This is an instance of the *it*+ADJ construction, where *it* constitutes the subject and the subclause an adjunct. Thus, the argument-to-function mapping for the two sentences where *it* and a dative experiencer co-occur is the one given in (42-b).

For the verb *gelimpan*, the most common pattern in the corpus is when *gelimpan* occurs with a subject *it*, with no dative experiencer. It is also common that *gelimpan* co-occurs with a propositional subclause only, with neither dative experiencer nor a subject *it*. This gives support for the argument structure in (43-b), where there is only one argument, the propositional argument, which is either associated with *it* or with the subclause. When *it* is present, it is this *it* that constitutes the subject, and the subclause constitutes an adjunct. When the subclause occurs on its own, it is the subclause that constitutes the subject.

Apart from these patterns, there is also a considerable number of sentences where *gelimpan* occurs with a dative experiencer instead of a subject *it*, giving support for the argument structure in (43-a), where *gelimpan* has the same argument structure as *thyncan*$_1$ in (42-a). For the verb *gelimpan*, as shown by the data here, there is a complementary distribution between a subject *it* and a dative experiencer. This can be taken as evidence for the argument structure where the dative experiencer is the SUBJ.

For the verb *thyncan*, there is no complementary distribution to be found between dative experiencers and *it* as can be seen in Table 5.3. There is thus no distributional support for the dative subject analysis.

On the other hand, there is support to be had from extraction. When a dative experiencer co-occurs with a *that*-clause, it seems to be possible to extract out of the *that*-clause. One example is given in (45).

(45) Hwæt þincð þe [þæt þu sy]?
what seems you.DAT that you be
'What do you think you are?'
(cowsgosp,Jn_[WSCp]:8.53.6483)

In (45), the *wh*-phrase *hwæt* is extracted out of the *that*-clause, giving support to the argument structure in (42-a), where the dative experiencer is SUBJ and the propositional argument COMP. If the subclause is a subject, we would not expect extraction to be possible here, as subjects typically are syntactic islands.

The argument structures given in (42) and (43) for *þyncan* and *gelimpan* are different from the raising verb analysis given for verbs such as *seem* and *appear* in (12) in Section 5.2.1. As discussed there, the *it*+COMP construction, where a non-thematic subject *it* occurs in conjunction with a complement subclause, is connected to the analysis of the predicate as a raising predicate. If the *it*+COMP construction were a part of Old English grammar, we would llso expect there to be raising predicates in Old English, alternately occurring in the raising construction and the *it*+subclause construction. This does not seem to be the case in the corpus material.

With respect to the verb *þyncan*, whose argument structures is given in (42), there are 15 instances in the Old English prose corpus, where this verb occurs together with an infinitival clause. In 14 out of 15 sentences, a thematic argument of the main clause is also the subject of the infinitival clause, i.e. it is control rather than raising that gives the identification of the subject of the infinitival clause. Consider the sentence in (46).

(46) Sumum menn wile þincan syllic [þis to gehyrenne],
Some.DAT men.DAT will seem strange this to hear
'To hear this must seem strange to some people.'
(coaelive,+ALS_[Maccabees]:564.5198)

Here, the subject of the infinitival clause is the dative experiencer *sumum men*, which also has a thematic role in relation to the predicate of the main clause. The phrase *sumum men* has an experiencer role in relation to the predicate *þyncan*. The structure in (46) is thus control rather than raising, as raising predicates take an arg1[-r] slot which is not assoicated with a thematic role.

There is, however, one sentence that has been analysed as raising in the literature (e.g. Denison, 1993: 221). This sentence is given in (47).

(47) swa þæt me þynceþ [of gemynde beon] Paulines
so that me.DAT seems out of memory be Paulinus'
wundor Nolane burge biscopes,
miracle Nola city bishop
'so it seems to me that the miracle of Pauline, bishop of the city of Nola, is forgotten'.
(cogregdC,GDPref_and_3_[C]:0.179.4.2177)

At a first glance, the nominative phrase *Paulines wundor Nolane burge biscopes* here seems to constitute the subject of both the verb *þyncan* and within the infinitival clause *of gemynde beon* (cf. Denison, 1993: 221). If we assume that *Paulines wundor Nolane burge biscopes* has no thematic role in relation to the verb *þyncan*, the sentence should then be analysed as raising, thus constituting a counterexample to the argument structures given in (42). However, considering the possibility of non-nominative subjects mentioned above, there is also an alternative analysis in which the dative pronoun *me* constitutes the subject of both *þyncan* and the bare infinitive *beon* and in which the phrase *Paulines wundor Nolane burge biscopes* constitues the copular complement of *beon*. Under such an analysis, the structure in (47) would be analysed as control and not raising. This is further supported by examples such as the one in (48), in which a dative pronoun seems to be the subject of the phrase *on gemynde beon* ('have in mind'). The example is taken from the *Charter of King Æthelbert to Sherborne* (Robertson, 1956: 16).

(48) Forþon ic Æþelbreht mid Godes gife Westsaxna
For-that-reason I Æthelbert with Gods gift Wessex
kyning witoðlice ic þence & me on gemynde is mid þissum
king truly I intend and me in mind is with these
earþlicum ðingum ecelican gestreon to begitanne.
earthly things everlasting treasure to procure
'For that reason, I, Æthelbert, king of Wessex, truly intend and have it in mind to procure the everlasting tresure by means of these earthly things.'

Furthermore, as discussed in Denison (1993), it is somewhat questionable whether the particular example given in (47) constitutes 'natural' Old

English. As pointed out by him (1993: 221), the structure and word order of the sentence seems to follow the Latin original in a rigid manner[12].

Thus, the lack of raising structures for the verb *þyncan* (cf. Elmer, 1983: 161) is an additional piece of support for the argument structures in (42), and for the fact that the *it*+subclause construction with the verbs *thyncan* and *gelimpan* should be analysed as *it*+ADJ, rather than *it*+COMP.

If we proceed to the Middle English and Early Modern English periods and consider the semantic equivalents there of *þyncan* and *gelimpan*, namely *seem* and *happen*, a different picture emerges.

In late Middle English we find raising constructions alternating with *it*+subclause constructions for the verb *seem*. This alternation is exemplified in (49) with sentences from Thomas Malory's *Morte d'Arthur* from about 1470.

(49) a. for he semed [to be ryght wyse].
(CMMALORY,34.1098)
b. Madam, hit semyth by your wordis [that ye know me].
(CMMALORY,658.4557)

In (49-a), we have an example of a raising construction with *seem*, and, in (49-b), from the same text, we also find an *it*+subclause construction. Based on a quantitative investigation in Gisborne & Holmes (2007), it is possible to conclude that the earliest attested examples of this alternation with *seem* can be found in the period 1350-1420 in the late Middle English period.

With respect to the verb *happen*, the same alternation is first attested in the Early Modern English period.

(50) a. And whan he happeneth [to rede or here any fable or historie], ...
'And when he happened to read or hear any fable or history'
(ELYOT-E1-P1,30.11)
b. in them it hapneth [that one in an other as moche deliteth as in him selfe].
'In them it happened that one delights in another as much as in himself.'
(ELYOT-E1-P1,161.180)

[12]The Latin original is *ita ut Paulini miraculum, Nolanae urbis episcopi, ..., memoriae defuisse videatur* (Denison, 1993: 221).

The sentences in (50) derive from Thomas Elyot's (1490-1546) *The boke named the Gouernour*. In the (49-a), we have a raising construction with the predicate *happen*, and, in (50-b), an *it*+subclause construction.

Thus, different verbs develop the argument structure, like the one in (12), with an arg1[–r] unassociated with a thematic role at different points in time (cf. Gisborne & Holmes, 2007). The way in which the development of raising verbs take place will not be further discussed here, but it is seems as if the verbs that become raising verbs take an experiencer role associated with arg1[–r], a experiencer role that at some point in time is reanalysed as being associated with a different argument slot (which is optional), leaving arg1[–r] unassociated with a thematic role (cf. Barron, 1997, 2001).

5.4 Summary

This chapter has discussed the data and analysis of the preposed clausal subject construction and the *it*+subclause construction in historical English, touching also upon issues concerning the null+subclause construction. The following conclusions have been reached. In Early and Late Modern English, infinitival clauses sometimes occur in the structural subject position, while *that*-clauses and *wh*-clauses do not. With respect to structural position, subject raising and coordinate subject deletion show positive evidence for the subject status of the subclause, while subject control and verb agreement are inconclusive as tests. For the *it*+subclause construction, there is evidence to support the conclusion that there are two separate constructions: (i) the *it*+COMP construction, with a non-thematic subject *it* and a complement subclause, and (ii) the *it*+ADJ construction, with a thematic subject *it* and an adjunct subclause. Furthermore, it was shown that the *it*+COMP construction occurs in connection with raising verbs, including *be*, while the *it*+ADJ construction occurs in connection with non-raising predicates.

Furthermore, for Old English, it was shown that the occurrences of preposed clausal subjects should probably be analysed as calques of their Latin originals, since there were no examples of preposed clausal subjects in non-Latin-based Old English. With respect to the *it*+subclause construction in Old English, it is posited that only the *it*+ADJ construction was available, and not the *it*+COMP construction. This conclusion is based on co-occurrence patterns, extraction data, and the non-existence of the raising construction with the predicates examined, *þyncan* and *gelimpan*. The first attested examples of the preposed clausal subject construction

and the *it*+COMP construction can be found during the Middle English period. The first attested example for a preposed *that*-clause subject is considerably later than the corresponding first examples for preposed *wh*-clause and infinitival clause subjects.

Part III

Weight, complexity and information structure

Chapter 6

Background

In part II of this dissertation, an analysis was given of the syntax and argument structure of predicates alternately taking the preposed clausal subject construction and the *it*+subclause construction. In this part, three factors thought to influence the choice of one alternative over another are discussed: (i) weight, (ii) complexity and (iii) information structure.

It has been known for a long time that both weight and information structure are important for the choice of alternative word orders. Behaghel (1909), for instance, presents the following two rules, or principles, supposed to govern word order:

(1) a. Das Unwichtigere (dem Hörer schon Bekannte) steht vor dem Wichtigen.
'Less important information (what is already known to the hearer) precedes the important information.'

b. Von zwei Satzgliedern geht, wenn möglich, das kürzere dem längeren voraus.
'Whenever possible, the shorter, out of two constitutents, precedes the longer.'

These two tendencies, i.e. that discourse-old content precedes discourse-new, and that light material precedes heavy material, are also included in modern reference grammars (cf. Huddleston & Pullum, 2002: 1372), sometimes associated with the terms end weight and end focus (e.g. Quirk et al., 1985: 1357, 1398).

Before going into the analysis and discussion of the present data in Chapter 7, it seems appropriate to present a selection of previous studies. The first section is concerned with previous studies on weight and complexity, while the second section focus on information structure.

6.1 Weight and complexity

Weight and complexity are two processing-related concepts that seem to have an important role to play with respect to word order. When applied to various linguistic phenomena, these concepts have received a wide array of definitions. Typical examples of defining criteria are the number of words or syllables of some domain, the number of nodes in a syntactic tree or layers of embedding. With respect to the alternation between the preposed clausal subject construction and the *it*+subclause construction, Jespersen (1909-1948) makes the following comment in his discussion on what he calls the preparatory *it*, i.e. the pronoun *it* in the *it*+subclause construction:

> *It* stands as a preliminary representative of a longish group of words which follows later because its placement here would make the sentence-structure cumbersome or 'top-heavy' (Jespersen, 1909-1949: part VII, 142).

In the above citation, Jespersen refers both to something that could be called weight ('longish group of words') and to something that could be called complexity ('cumbersome or top-heavy' structure). According to this quote, a group of words is extraposed because it would otherwise make the sentence hard to process as a result of a 'cumbersome or top-heavy' structure.

In more recent studies, various models of efficient sentence processing have been proposed trying to capture how weight and complexity can, and ought to be, operationalised. One of these models is presented in Hawkins (1994, 2004).

In this section, Hawkins' model is presented, followed by a short presentation of a study making use of his methodology to discuss relative clause extraposition, Francis (2010). Lastly, a short presentation is given of a study not making use of Hawkins' methodology, but nonetheless focusing on the influence of weight on the preposed clausal subject construction, Erdmann (1988).

6.1.1 Hawkins' processing model

Hawkins (1994, 2004) proposes three principles facilitating the processing of sentences: (i) minimising the sequences required to recognise combinatorial and dependency relations, (ii) minimising the formal complexity of each linguistic form, and (iii) maximising the number of properties that can be assigned early in a sentence (Hawkins, 2004: 31-61). These

principles are termed (i) the Minimize Domains, (ii) the Minimize Forms and (iii) the Maximize On-Line Processing. The principles are given in (2).

(2) a. Minimize Domains:
The human processor prefers to minimize the connected sequences of linguistic forms and their conventionally associated syntactic and semantic properties in which relations of combination and/or dependency are processed. The degree of this preference is proportional to the number of relations whose domains can be minimized in competing sequences or structures, and to the extent of the minimization dispreference in each domain. (Hawkins 2004: 104)

b. Minimize Forms:
The human processor prefers to minimize the formal complexity of each linguistic form F (its phoneme, morpheme, word, or phrasal units) and the number of forms with unique conventionalized property assignments, thereby assigning more properties to fewer forms. These minimizations apply in proportion to the ease with which a given property P can be assigned in processing to a given F. (Hawkins, 2004: 38)

c. Maximize On-Line Processing:
The human processor prefers to maximize the set of properties that are assigned to each item X as X is processed, thereby increasing O(nline) P(roperty) to U(ltimate) P(roperty) ratios. The maximization difference between competing orders and structures will be a function of the number of properties that are assigned or misassigned to X in a structure/ sequence S, compared with the number in an alternative. (Hawkins, 2004:51)

Hawkins operationalises the efficiency principles in the form of two different ratios: (i) the Immediate Constituents to words (IC-to-word) ratio and (ii) the On-line Properties to Ultimate Properties (OP-to-UP) ratio. For the purposes of this study, the IC-to-word ratio will have to be further explained, as this ratio is applied to my data in the next chapter.

Introduced by Hawkins (1994), the IC-to-word ratio concerns the number of words it takes for an interpreter to recognise all the immediate constituents (ICs) of some domain, for example a main clause or a particular phrase. This ratio is calculated as the number of ICs divided by the number of words, which, in its turn, can be translated into a percentage.

The higher the percentage, the fewer the words required to recognise the ICs, thus making the sentence more efficient to process and less complex. It follows that the measurement is focused on comprehension rather than production. However, considering the fact that there is an interrelation between production and comprehension in both language acquisition and language use in general, it is reasonable to suppose that the ratio, if focused on comprehension, also indirectly pertains to production.

In order to illustrate the use of this ratio, consider the following two sentences.

(3) a. . . . [[W]hether she plays well] appears to be a matter of chance. [BNC]
b. It appears to be a matter of chance [whether she plays well]. [constructed]

If we take the domain to be the main clause, the preposed clausal subject construction in (3-a) has two immediate constituents: (i) the subject *whether*-clause and (ii) the predicate (the finite verb and the complement infinitival clause). The *it*+subclause construction in (3-b) has three immediate constituents, which in addition to the constituents of the preposed clausal subject construction also contains a subject *it*.

For (3-a), there is an IC-to-word ratio of 2/5=40%. It takes five words for the two relevant constituents to be recognised. The constituent *whether she plays well* contains four words and then, by reaching the finite verb, the constituency is made clear. For (3-b), the ratio is 3/9=33%, i.e. nine words are required for the recognition of three constituents. According to this quantitative approach, there would thus be a slight processing advantage for the preposed clausal subject construction in (3-a). As it happens, the excerpted sentence is indeed a preposed clausal subject construction in the corpus. However, it must be pointed out that the differences in ratio here are very slight. Therefore, it cannot be concluded that the differences in complexity account for the constructional choice in this particular case. However, if the differences in weight distribution become larger, as in (4) or (5), we can see that the preference for one construction over another becomes clearer.

(4) a. It is disputed [whether these onion domes were a development indigenous to the area or whether the idea came from further east.] [BNC]
b. [Whether these onion domes were a development indigenous to the area or whether the idea came from further east] is disputed. [constructed]

(5) a. [Whether that occurs] depends on its responses to the issue that is going to continue to dominate the political scene – the economy in general and the consequences of Exchange Rate Mechanism membership in particular. [BNC]
b. It depends on its responses to the issue that is going to continue to dominate the political scene – the economy in general and the consequences of Exchange Rate Mechanism membership in particular – [whether that occurs]. [constructed]

In (4-a), three immediate constituents are recognised in four words, which gives an IC-to-word ratio of 3/4=75%. In (4-b), on the other hand, 21 words are required to recognise the two immediate constituents. The IC-to-word ratio of (4-b) is thus 2/21= 10%. There is thus a clear processing advantage for the attested extraposition construction in (4-a).

In (5) the situation is reversed. Here there is a clear preference for the attested preposed clausal subject construction in (5-a). For this construction, five words are needed to recognise two immediate constituents, representing an IC-to-word ratio of 2/5=40%. For the *it*+subclause construction in (5-b), 33 words are needed to recognise three constituents, yielding an IC-to-word ratio of 3/33=9%. Thus, in both (4) and (5), the attested sentences are the ones with the most favourable ratios.

Hawkins' principles for efficient sentence processing have been tested on a number of occasions in previous research (e.g. Uszkoreit et al, 1998; Konieczny, 2000). In the next section, we will take a look at one study that is of particular relevance for the current investigation, namely Francis (2010).

6.1.2 Hawkins' model tested

Francis (2010) tests Hawkin's domain minimisation principles in a series of psycholinguistic experiments on relative clause extraposition (RCE). RCE entails that the relative clause is detached from its accompanying noun and is placed in the rightmost position in the clause. The example in (6) shows first the relative clause extraposition construction and then the corresponding nonextraposition construction where the relative clause immediately follows the head noun.

(6) a. New sets soon appeared that were able to receive all the TV channels. (ICE-GB)
b. New sets that were able to receive all the TV channels soon appeared.

In order to test the extent to which Hawkins' model can account for the alternation exemplified in (6), Francis (2010) performed three experiments: (i) a reading time test, (ii) an acceptability judgement task, and (iii) a corpus investigation.

In the reading time experiment, forty informants were asked to read sentences with the two structures in (6), where the length of the relative clause varied between 4, 8 and 15 words, but where the length of the VP was held constant at 5 words. Based on measurements of the mean reading time per word, Francis concluded that there was a significant reading time advantage for extraposition over the nonextraposition structure when the relative clause was heavy (15 words long). When the relative clause was light (4 words long), there were no significant differences.

In the acceptability judgement task, 31 informants were given a survey in which they were asked to rate the same sentences as those in the reading time experiment. In line with the predictions in Hawkins' model, acceptability ratings for nonextraposition sentences decreased as the weight of the relative clause increased. When the relative clause was light (four words long), there was a slight advantage for the nonextraposition structure.

The corpus investigation was based on material from the International Corpus of English, specifically the British part (ICE-GB). In the coding, Francis (2010) looked at the type of main predicate, type of relative clause (subject, direct object, object of preposition, possessive, or adjunct) and type of discourse (spoken or written). In the corpus, 391 sentences were collected. Out of these, 59 (15%) had an extraposed relative clause and 332 (85%) had the nonextraposition structure. On average, it turned out that extraposed relative clauses were longer than the VP, while nonextraposed RCs were shorter. Finally, the proportion of sentences with RCE decreased as the ratio of VP length to RC length increased (Francis, 2010).

An interesting thing to note in all of Francis' (2010) experiments is that extraposition seems to require a considerable advantage in terms of the weight distribution. When the differences are not that great, nonextraposition seems to be preferred. Thus, in the case of relative clauses, it seems as if there is a clear advantage for the nonextraposition structure when the factor of weight is out of play. In those cases where the RC and VP were exactly the same length, extraposition occurred in only 9% of the sentences even though the IC-to-word ratio would have predicted them to occur at about the same frequency.

6.1.3 Erdmann (1988) on weight and subject extraposition

One relevant study about weight in connection to subject *it*-extraposition is that of Erdmann (1988). In his study, he compares the weight of main clause adjectival predicates in *it*+subclause constructions as opposed to preposed clausal subject constructions. His data consist of 452 occurrences of subject *that*-clauses, where 416 are extraposed and 36 nonextraposed. In contrast to studies where weight is treated as continuous, Erdmann treats it as a categorical variable. The notion of weight for Erdmann is based on whether or not the adjective in the superordinate clause is postmodified or coordinated with another adjective. Thus, if the adjective is coordinated or postmodified, it is heavy; otherwise it is light. Strangely enough, premodification is not taken as a indicator of heaviness, a decision which is not explained. The following sentences from Erdmann (1988) exemplify light and heavy predicates according to his classification:

(7) a. It is *true* that .... [light]
b. It is *exciting and surprising* that .... [heavy]
c. It was *less self-evident than he thought* that .... [heavy]
d. It is *important, however,* that .... [heavy]

The results obtained by Erdmann are presented in Table 6.1, which shows my representation and statistical analysis of Erdmann's data.

Table 6.1: Weight of the main clause predicate in Erdmann (1988) as a function of constructional choice

	Light	Heavy	Total
Preposed clausal subject	20	16	36
it+subclause	365	51	416
Total	385	67	452

(Fisher's exact test, p-value < 0.05, odds ratio $= 0.18$)

Table 6.1 indicates that 365 (88 %) occurrences of the *it*+subclause construction have a light predicative adjective, and that 51 (12 %) have a heavy one. For the preposed clausal subject construction, on the other hand, 16 (44 %) predicative adjectives are light and 20 (56 %) heavy (Erdmann, 1988: 330). Applying a Fisher's exact test to Erdmann's data shows that the distribution represents highly significant differences between the preposed clausal subject construction and the *it*+subclause

construction with regards to the weight of the predicative adjective. The results leads Erdmann to formulate the principle of balanced weight: 'if the superordinate predicative adjective is light, the heavy clausal subject tends to be extraposed' (1988: 33).

Apart from Erdmann (1988), there are a number of other studies that are not specifically concerned with subject extraposition but with other related word order alternations such as Heavy NP Shift, Particle Shift, and Dative Shift. Among the relevant studies we find Wasow (1997), Arnold et al. (2000), Lohse et al. (2004) and Stallings & MacDonald (2011). While not going into any detail with regard to these studies, we can at least summarise them by stating that weight seems to be highly relevant for the word order alternations mentioned above. Further, in terms of what measurements of weight are most revealing, Wasow (1997) argues that, generally speaking, there are no significant differences between the different ways of measuring weight. As will be seen from my study later on, Wasow's conclusion about different measurements of weight can be partly supported, although not in full.

6.1.4 The IC-to-word ratio vs. relative weight

Before concluding the background treatment of weight and complexity, it seems relevant to discuss the fact that there is a certain correlation between Hawkins' IC-to-word ratio and a straightforward relative weight measurement.

One case where differences between these methods can be discerned concerns preposed clausal subject constructions embedded within a subclause. A subclause embedded within another clause with material on both sides of it, is commonly known as center-embedding. Center-embedding in general seems to be bad in terms of sentence processing (cf. Karlsson, 2007). Consider example (8), where the nonextraposed *whether*-clause in (8-b) is center-embedded. The *whether*-clause in (8-a) occurs in a clause-final extraposed position and is consequently not subject to the same processing difficulties. The relevant constructions are given in italics.

(8) a. An interesting feature of the general theory of relativity is that *it should not matter [whether time is running backward or forward]*: [BNC]

b. An interesting feature of the general theory of relativity is that *[whether time is running backward or forward] should not matter*: [constructed]

Within the subclause, as can be calculated, there is a considerably better IC-to-word ratio for the *it*+subclause construction in (8-a). All relevant constituents (subject and predicate) will have been detected by the time the interpreter reaches the word *whether*, which is the fifth word. For the preposed clausal subject construction in (8-b), on the other hand, the interpreter needs to get to the eighth word *should* before the subject and the predicate can be recognised. Based on the ratios for the alternatives in (8), the use of extraposition is predicted. However, consider what happens if we embed the *whether*-clause within a nominal phrase.

(9) An interesting feature of the general theory of relativity is that *the question of [whether time is running backward or forward] should not matter*: [constructed]

In (9), the subject is made longer by three words. However, as can be seen, the *whether*-clause is no longer center-embedded, but rather contained in a complex NP, and the utterance as such thus becomes more acceptable than the center-embedding construction in (8-b). This fact does not follow immediately from Hawkins' metrics as presented in the previous section.

What is missing from Hawkin's metrics is the ban against centerembedding having something to do with the ease at which an interpreter can recognise the *whether*-clause as a subject. For both (8-b) and (9) the interpreter cannot be certain about the grammatical function of the initial element, the clause or the complex NP, until the finite verb is reached. The reason why (9) is considerably more acceptable could then be tied to the probability of the relevant constituent being a subject. While an initial *whether*-clause could be an adjunct or a fronted complement, an initial nominal phrase is, at least statistically speaking, very likely to be a subject. Thus, even though the nominal subject is longer than the corresponding clausal subject, the sentence in (9) is still more acceptable as it is easier for the interpreter to recognise its nominal constituent as a subject. We can thus see that the likelihood of a constituent having a certain function seems to play an important role in the processing of the above sentences, something which is not really captured by Hawkins' metrics. Of course, the claim that the probability at which a phrase is interpreted as a subject affects the acceptability of that phrase as a subject does not provide an explanation as to why a phrase does not occur in a certain position. Part of the explanation might also be found in certain prosodic constraints and the need for a subordinate clause to forms its own intonational phrase (cf. Light, 2012: 164). At any rate, the

relationship between prosody and extraposition is something that needs to be further explored in future research.

In short, the present section has given a presentation of previous studies on the concepts of weight and complexity in relation to the choice between the preposed clausal subject and the *it*+subclause construction. In the next section, studies will be discussed concerning the same syntactic alternation, but now with a focus on information structure.

6.2 Information structure

The information structural differences between the preposed clausal subject construction and the *it*+subclause construction have been discussed on a number of occasions in the literature. In this section, three studies of particular relevance are presented and discussed: Miller (2001), Birner & Ward (2004) and Bolinger (1977). The primary aim of these studies can be described as a search for the necessary and sufficient information structural conditions for the use of the preposed clausal subject construction. Furthermore, it should be noted that none of them deals with the phenomenon of weight and complexity in relation to information structure.

The reason why the above-mentioned studies focus on the preposed clausal subject construction and not on the *it*+subclause construction is probably a result of the fact that the *it*+subclause construction is often seen as the 'unmarked' alternative in relation to the preposed clausal subject construction (e.g. Biber et al., 1999: 676). While I am hesitant to use the term unmarked, previous studies (e.g. Kaltenböck, 2005) nonetheless suggest that the *it*+subclause construction is less pragmatically restricted and more diversified in comparison to the preposed clausal subject construction. The studies presented in this chapter will primarily concern the preposed clausal subject construction.

6.2.1 Miller (2001)

Miller (2001), based on an examination of a corpus of naturally occurring examples, provides an analysis of the discourse condition on the alternation between the preposed clausal subject construction and the *it*+subclause construction. To begin with, he (2001: 687) claims that

> [...] the reference of the sentential or infinitival VP subject must be discourse-old or directly inferable from the previous discourse context in order to remain in subject position. If this condition does not hold, extraposition is obligatory.

Miller's claim is backed up by the judgements of alternative pairs such as the one in (10), which constitutes the opening of a newspaper article.

(10) a. European Central Bank Row Won't be Last
PARIS It is astonishing [that the real questions about Europe's new single currency, the euro, and about the new European Central Bank were never addressed during the 12-hour row among European governments that ended in Sunday's sad compromise on the new bank's president]. (Miller, 2001: 690)

b. #[That the real questions about Europe's new single currency, the euro, and about the new European Central Bank were never addressed during the 12-hour row among European governments that ended in Sunday's sad compromise on the new bank's president] is astonishing.

The example in (10), on Miller's account, is meant to show that the preposed clausal subject in (10-b) is unacceptable because it is not discourse-old or directly inferable from the context. As the sentence constitutes the opening of a newspaper article, it is argued that the content of the subject clause cannot have been mentioned previously or be derived from the previous discourse.

Looking at Miller's example, one might wonder whether the fact that it appears as the opening of a newspaper article is the only thing that makes this sentence unacceptable. One thing that is striking about the example is its weight distribution: the *that*-clause is almost 20 times longer than the main clause predicate. Without taking weight into account and considering sentences with different weight distributions, it seems difficult to base the information structural conditions on preposed clausal subjects on sentences such as the one in (10-b) alone. As we saw in chapter 6, there is a strong pressure in terms of weight and complexity for a sentences with a weight distribution like the one in (10) to be realised as *it*-extraposition. The relationship between information structure and weight/complexity is thus something which needs closer examination, and the issue will be further discussed in relation to my own data in Section 7.4.

6.2.2 Ward & Birner (2004)

Another study that uses the concept of givenness in connection with extraposition is Ward & Birner (2004: 167-168). Generally speaking, they agree with Miller that the nonextraposition of a content clause requires certain conditions on the givenness of that clause. In contrast to Miller,

however, they hold that it is the hearer-status rather than the discourse-status that is important. Preposed subject clauses, it is claimed, need to be hearer-old, i.e. assumed to be known by the hearer, but not necessarily discourse-old, i.e. evoked in the previous discourse. In support of their claim, they refer to the acceptability of the following example.

(11) His act takes on lunatic proportions as he challenges female audience members to wrestling matches, falling in love with one while grappling it out on the canvas. [How he and feminist Lynne Margulies (Courtney Love) became life partners] is anyone's guess. (Ward & Birner, 2004: 167)

Apparently, the content of the *how*-clause in (11) is not evoked in the previous discourse. Yet, according to Ward & Birner's analysis, it could still be deemed acceptable, as long as the speaker assumes that the hearer knows that the referents of *he* (Kurt Cobain) and Lynne Margulies are life partners. The nonextraposition of the *how*-clause is thus supposed to give rise to the presupposition that the speaker treats the *how*-clause as hearer-old.

The felicity of (11) supports the hypothesis that being hearer-old is relevant for the felicity of preposed subject *wh*-clauses. However, it is not obvious that examples such as the present one would be equally good with other types of clauses, e.g. infinitival clauses or declarative content clauses. First, a *that*-clause typically refers to a proposition, while a *wh-clause* tends to refer to a question. Second, the word *that* is a pure subordinator (e.g. Huddleston & Pullum, 2002), while the *wh*-element participates in the clause as an argument or an adjunct. These differences make it difficult to use a *wh*-clause to argue against, for instance, Miller's claims, which are based on the typical behavior of *that*-clauses. Thus it is possible that *that*-clauses, infinitival clauses and *wh*-clauses are subject to different discourse constraints in relation to extraposition. As the sentence in (11) has a more even weight distribution than the sentence in (10), it might be the case that this difference in weight distribution affects the acceptability, as argued already in Section 6.1.

6.2.3 Bolinger (1977)

A third study about givenness and extraposition is Bolinger (1977). He analyses the *it*-pronoun of the *it*+subclause construction as a referential anaphor, and not as a dummy pronoun devoid of meaning. The consequence of this analysis, according to Bolinger's hypothesis, is that the *it*-pronoun requires a contextually established referent that it could refer

to. Since the subject *it* and the subclause are coreferential, the claim is thus that *it*-extraposition is used when the subclause contains familiar information. Consider the example in (12), taken from Bolinger (1977: 73).

(12) What do you think of running him as a candidate?
  a. *[To do that] would be a good idea.
  b. It would be a good idea [to do that].

Bolinger claims that the speaker in (12) is required to use *it*-extraposition, as in (12-b), and that the nonextraposition construction in (12-a) is unacceptable. The fact that the speaker uses the phrase *to do that*, thus picking up the content of the VP *running him as a candidate*, means, according to Bolinger, that *it*-extraposition is obligatory.

It needs to be pointed out that the dialogue presented in (12) seems slightly problematic whatever structure is chosen in the response sentence. The most natural response to the question in (12) would probably be to use a pronoun to refer back to *running him as a candidate*, for instance *that*, rather than the entire infinitival phrase *to do that*, i.e. *that would be a good idea*. The less pragmatically restricted *it*-extraposition structure seems to be possible, whereas the preposed clausal subject construction seems odd. The fact that the sentence in (12-a) is odd provides problems for the claims of both Miller (2001) and Birner & Ward (2004), where it would be predicted to be fine by both accounts.

Arguably, the theory presented by Bolinger has a number of problems. One problem arises when we consider a number of naturally occurring examples cited by Miller (2001). Miller, who, as we saw above, claims that a nonextraposed clause is required to be discourse-old, provides a number of examples of non-extraposed infinitival clauses containing anaphoric elements. Let us look at some of Miller's (2001: 686) examples and consider them in relation to the claims put forward by Bolinger concerning the example in (12-a).

(13) a. "So you get rid of that pistol right now, Mister McBride. You do that or take you out a permit right now." McBride couldn't do either, of course. Not immediately, as the deputy demanded. Not without a face-saving respite of at least a few minutes. [To do so] would make his job well-nigh impossible.
  b. His [Faulkner's] denials of extensive reading notwithstanding, it is no doubt safe to assume that he has spent time schooling himself in Southern history and that he has gained some

acquaintance with the chief literary authors who have lived in the South or have written about the South. [To believe otherwise] would be unrealistic.

c. Neither had a choice other than to accept the invitation. [To have refused] would have been political suicide.

In (13), all three clausal subjects in the nonextraposition constructions contain anaphoric elements. In (13-a), *to do so* refers back to *do either* a couple of sentences before; in (13-b), *otherwise* links back to a previously expressed belief; and, in (13-c), *refused* points back to *accept* in the immediately preceding sentence. Miller's examples thus seem to disprove Bolinger's claim that extraposition is obligatory when there is a prior basis established for the content of the clausal subject.

However, although Bolinger's analysis seems to be questionable, it does seem as if the acceptability judgements in (12) are correct. When considering the examples given by Bolinger (1977) and Miller (2001), there is one thing that distinguishes Miller's examples from Bolinger's (12-a), namely CONTRAST. Miller's examples in (13) all represent contexts in which there is a clear contrastive relation between the content of the preposed clausal subject and some content in the preceding discourse. In (13-a), *to do so* contrasts with *not doing so*; in (13-b), *to believe otherwise* contrastively refers back to the particular belief expressed in the previous discourse; and, in (13-c), *to have refused* contrasts with accepting the invitation, which is mentioned explicitly in the immediately preceding sentence. In the one previous example of an acceptable preposed clausal subject in the present section, the sentence in (11), there is likewise a clear contrastive interpretation.

In Bolinger's (12-a), on the other hand, there is nothing in the preceding context that *to do that* can contrast with. It seems to be this lack of a possible contrastive context that makes Bolinger's example unacceptable. Interestingly, with a contrastive context for (12-a), the acceptability is considerably improved. Consider the constructed example in (14).

(14) A: I'm not sure whether we should run him as a candidate or not. What do you think?
B: [To do that] would be a really good idea. No doubt about it.

With a contrastive context, where there is a polar contrast between doing that and not doing that, example (14) is considerably better than (12-a).

The examples given by Miller do thus not invalidate the acceptability judgements made by Bolinger. They do however call into question his analysis of these sentences. Instead of givenness per se, it is the presupposition

of contrast that seems to have an important effect on the acceptability of preposed infinitival clausal subjects. The phenomenon of contrast is not discussed by Bolinger (1977) or Miller (2001) in connection to preposed clausal subjects. However, based on the examples they give, it seems to be of utmost importance. Furthermore, it is not surprising that contrast should be relevant for preposed clausal subjects, as it corresponds to the general properties of preposing in English, as described in Birner & Ward (1998).

Epitomising the present section, we have seen that, among the previous studies discussed, the information structure of the preposed clausal subject construction has typically been described in terms of constraints on givenness. Miller (2001) argues that the subclause in the preposed clausal subject construction needs to be discourse-old or directly inferrable, while Ward & Birner (2004) claim that the subclause needs to be hearer-old. In the discussion of the examples used in the previous studies, it seems that one important phenomenon in relation to the use of nonextraposition, perhaps overlooked in these studies, is contrast. It was proposed that certain examples might be better explained in terms of a constraint on contrast, rather than one on givenness. Furthermore, it was pointed out that the phenomena of weight and complexity need to be discussed in relation to the information structural properties of the constructions. On the whole, it seems difficult to argue for the acceptability of various options without taking into account weight and complexity.

6.3 Summary

This chapter has discussed previous studies on weight, complexity and information structure in relation to word order alternations in general, and the preposed clausal subject construction in particular. With respect to weight and complexity, Hawkins' (2004) proposed model was first presented, followed by one study testing the predictions of the model. The conclusion of the different studies seem to be that weight and complexity have a considerable effect on the choice of construction in word order alternations. With respect to information structure, the status of the subclause as discourse- or hearer-old has been proposed as the primary factor in previous studies. However, as discussed, it emerges from the examples given in these studies that the presence of a contrastive relation should not be overlooked.

Chapter 7

Results and analysis

As we have seen, Chapter 6 concerned previous studies on the weight, complexity and information structure of preposed clausal subjects and the *it+subclause* construction. In the present chapter, my own treatment of these phenomena is presented, including both a presentation of relevant corpus data and an analysis of the different findings.

Section 7.1 concerns the material used for the study, which not only includes the historical corpora discussed in Part II, but also the BNC sample described in Section 3.1.5. In Section 7.2, the results of a quantitative investigation of weight and complexity are presented, followed by the results of a corresponding investigation into information structure in Section 7.3. Finally, in Section 7.4, results concerning weight, complexity and information structure are brought together, and the different relationships between them are discussed.

7.1 Material

The material used in this chapter comes from two sources: (i) the Penn Corpora of Historical English and (ii) the British National Corpus (BNC). A description of the corpora has previously been given in Chapter 3.

For the material from the historical corpora, an analysis of weight, complexity and information structure was performed of the instances of the preposed clausal subject construction and the *it*+subclause construction. The null+subclause construction will not be included in the discussion in this chapter. The frequency of the two constructions in the historical corpora can be seen in Table 7.1.

Table 7.1: Distribution of the extraposition alternation in the historical corpora.

	OE	ME	EME	LME	total
Preposed clausal subject	4	59	171	193	427
it+subclause	705	1392	3334	2151	7582

The table shows that the *it*+subclause construction is considerably more frequent in token numbers than the preposed clausal subject construction in all historical corpora. The material from the Old English period will not be discussed quantitatively here, as there are only four examples tagged as clausal subjects in the corpus. More information about the coding queries used to extract the relevant constructions can be found in Chapter 3.

As for Present-day English, the BNC sample, presented Section 3.1.5, is used. The sample contains 1,000 instances of the word *whether*. Out of these instances, 48 sentences contain preposed clausal subjects and 60 feature the *it*+subclause construction. Thus, it is these 108 instances that provide the basis for the discussion of weight, complexity and information structure in Present-day English in this chapter.

The fact that only *whether*-clauses are examined for Present-day English, while *that*-clauses, *wh*-clauses (including *whether*-clauses), and infinitival clauses are examined in the historical corpora, is something that needs to be taken into account when considering the results.

7.2 Weight and complexity

As described in Section 6.1, weight and complexity have been taken to be important underlying factors for the choice between the preposed clausal subject and the *it*+subclause construction. In the present section, the quantitative results relating to weight and complexity are discussed.

The section is organised as follows. First, the treatment of the notions of weight and complexity is presented. Then, the quantitative results for weight and complexity in the BNC sample is dealt with. Lastly, a discussion is maintained on the corresponding results of weight and complexity in the historical corpora.

7.2.1 Operationalisation of weight and complexity

In dealing with weight and complexity, three different measures are used. Apart from the IC-to-word ratio, which was described in Section 6.1.1, there are two different measurements of relative weight.

The first measure, which was applied manually to the BNC sample, involves dividing the difference between the number of words (clitics not counted) of (a) the subclause and (b) the predicate with the sum of these two values. The predicate is separated from potential adjuncts and the subject *it* is not included.

$$\frac{a-b}{a+b} = \text{relative weight value}$$

When the subject clause and the predicate have the same number of words, i.e. have the same weight, the relative weight value is ±0. A relative weight value of more than zero tells us that the subject clause is heavier than the predicate and vice versa if the the relative weight value is below zero. Further, the relative weight value +1 tells us that the subject clause constitutes 100 % of the construction, and the value –1 that the predicate constitutes 100 % of the construction. Neither the value +1 nor the value –1 are found in the sample as the sentences there always consist of a subject and a predicate. The sentences in (1) represent three different relative weight values. The first two sentences represent the preposed clausal subject construction, and the third sentence represents the *it*+subclause construction.

(1) a. But [whether it's therapeutic for the reader] is not for me to say. [relweight=±0, BNC]
b. But [whether she plays well] appears to be a matter of chance. [relweight=–0.27, BNC]
c. It is doubtful [whether this index provides an appropriate basis for measuring the rate of inflation]. [relweight=+0.73, BNC]

In (1-a), the subject *whether*-clause and the predicate are of the same weight (six words each), which gives a relative weight value of ±0. In (1-b), the predicate (with seven words) is heavier than the *whether*-clause (which has four words), and the relative weight value is –0.27. Lastly, in (1-c), the *whether*-clause (with 13 words) is heavier than the predicate (which has two words), and the relative weight value is +0.73.

In using the above calculation, the current study deviates from some earlier studies using subtraction to calculate relative weight (e.g. Arnold et al., 2000: 35). If the range of the data is large, as is the case here, subtraction is likely to give an unreliable result. When two sentences differ drastically in total number of words, the subtraction calculation gives a less accurate relative weight value in comparison to a division calculation. As a case in point, the two relative weight calculations 10-3=7 and 25-18=7 give the same relative weight value (=7) if subtraction is used. However, using the division calculation, the resulting relative weight values differ substantially (cf. 10–3)/(10+3)=0.54 and (25–18)/(25+18)=0.16). It is reasonable to assume that a difference of 7 words in length gives a different effect if the two values are 10 and 3 compared to 25 and 18. Hence, the division calculation gives a more reliable result than a study using subtraction.

The second measurement is a simplified version of the first one, and is applied to the historical material. It is based on a calculation of the length (in number of words) of the subclause divided by the length of the whole sentence. This measurement thus gives us a relative weight value, which says what part the subclause constitutes in relation to the sentence. In the corpus annotation for the historical corpora, there is no separation between arguments and adjuncts, which means that the predicate cannot be separated from adjuncts or from a subject *it*. The fact that there is no separation of the predicate from potential adjunct, and the fact that the subject *it* is counted as part of the sentence means that this measurement does not target the relative weight of the relevant constitutents as correctly as the first measurement. The coding queries used to get the number of words of the subclause and of the sentence are given in the appendix.

7.2.2 Weight and complexity in Present-day English

Featuring 108 instances, the BNC sample was checked for both IC-to-word ratios and relative weight values. As will be recalled from Section 6.1, the IC-to-word ratio concerns the number of words required for an interpreter to recognise all the immediate constituents (ICs) of some domain, for example a main clause or a particular phrase. Relative weight, on the other hand, concerns the number of words of the subclause in relation to the number of words of the predicate or the sentence as a whole. The results show that both these values, which are closely related, have a significant effect on the choice of construction.

The distribution of IC-to-word values for the *it*+subclause construction and the preposed clausal subject construction is shown in Table 7.2.

Table 7.2: IC-to-word ratio (number of words)

	Mean	SD	Sample size
Preposed clausal subject	0.19	0.08	48
it+subclause	0.57	0.14	60
(Wilcoxon rank sum test, W = 2844.5, p-value < 0.05)			

As can be seen, the mean IC-to-word value for the preposed clausal subject construction, which is around 0.2, is significantly lower than the mean value for the *it*+subclause construction, which is about 0.6. Thus, it generally takes more words to recognise the immediate constitutents in the case of the preposed clausal subject construction in comparison to the *it*+subclause construction. In fact, in 100 out of 108 occurrences, there is a lower IC-to-word value for the *it*+subclause construction. However, the *it*+subclause construction is realised in only 60 out of the 108 occurrences. If the IC-to-word value was the sole decisive factor in the choice between the *it*+subclause construction and the preposed clausal subject construction, we would expect to see considerably more cases of the *it*+subclause construction than we have actually found.

Turning to the distribution of relative weight values, the relevant data are presented in Table 7.3.

Table 7.3: Relative weight (number of words)

	Mean	SD	Sample size
Preposed clausal subject	0.30	0.40	48
it+subclause	0.58	0.21	60
(Wilcoxon rank sum test, W = 2073.5, p-value < 0.05)			

As can be seen, the relative weight of the predicate in relation to the subject is considerably higher for the *it*+subclause construction. Just like for Table 7.2, the difference between the two constructions in Table 7.3 is statistically significant.

A visualisation of the influence of relative weight on the choice of construction is given in Figure 7.1.

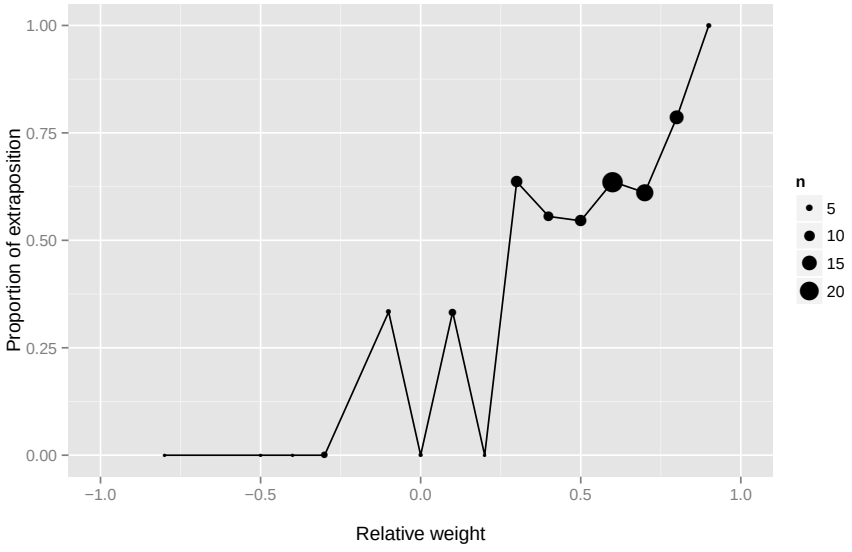

Figure 7.1: Proportion of extraposition in relation to relative weight

The figure[1] shows that, for the low values of relative weight, where the subclause is shorter in relation to the predicate, there is a greater proportion of preposed clausal subjects. There are 20 instances with a relative weight value between –0.8 and +0.2. Three of these are *it*+subclause constructions, while the rest, 17 constructions, are preposed clausal subject constructions. Thus, a low relative weight value seems to go hand in hand with the preposed clausal subject construction. The higher the relative weight value is, the greater the proportion of *it*+subclause constructions. For the values between +0.3 and +0.9, 57 occurrences represent the *it*+subclause construction, whereas 31 occurrences represent the preposed clausal subject construction.

An interesting observation emerging from Figure 7.1 is that there seems to be a tipping point in terms of relative weight at about +0.3. Above this value, the *it*+subclause construction dominates, while, below it, there is a majority of preposed clausal subjects. For the purpose of illustration, an example of a sentence with a relative weight of +0.3 is given in (2).

[1]For purposes of exposition in Figure 7.1, the relative weight values have been rounded off to one decimal point.

(2) [Whether disclosure is required if the Calcutta office bills the US parent direct] will depend on who requisitions the work. [BNC]

Here, as can be seen, the subclause contains 13 words and the predicate 7 words, yielding a relative weight value of +0.3 (13–7/13+7=0.3).

7.2.3 Weight and complexity in historical English

As mentioned above, the relative weight measure used for the historical data involves dividing the length of the subclause, with the length of the sentence, both calculated in terms of the number of words involved. The results of this calculation are in Table 7.4, with the first part representing Middle English, the second Early Modern English and the third Late Modern English.

Table 7.4: Relative weight (clause/sentence) in the historical corpora as function of constructional choice

Middle English	Mean	SD	Sample size
Preposed clausal subject	0.41	0.18	54
it+subclause	0.58	0.21	1339
(Wilcoxon rank sum test, W = 20498.5, p-value < 0.05)			

Early Modern English	Mean	SD	Sample size
Preposed clausal subject	0.41	0.20	160
it+subclause	0.57	0.24	2568
(Wilcoxon rank sum test, W = 123365.5, p-value <0.05)			

Late Modern English	Mean	SD	Sample size
Preposed clausal subject	0.47	0.21	190
it+subclause	0.58	0.22	1780
(Wilcoxon rank sum test, W = 120071, p-value < 0.05)			

The totals not being the same here as in Table 7.1 is a consequence of the fact that a limited number of sentences have not been coded for weight and givenness, which in turn is a consequence of the formulation of the coding query. To give one example of an instance that has not been coded for weight, consider the sentence in (3).

(3) [To do thus in Court], is counted of some, the chief and greatest grace of all:
(ASCH-E1-P2,14V.104)

The sentence in (3) constitutes the passive counterpart to the sentence in (4), which contains a small clause.

(4) Some count [to do thus in Court the chief and greatest grace of all]:
[constructed]

The sentence in (3) is a preposed clausal subject construction, and the sentence as a whole has been coded for weight and givenness. However, the small clause *to do thus in court (is) the chief and greatest grace of all*, which also has *to do thus in court* as its subject, has not been counted. The number of words of this subject is given as zero, meaning that this represents an instance coded as a preposed clausal subject construction, which has no relative weight.

The general impression from Table 7.4 is that relative weight seems to play an important role in the choice between preposed clausal subjects and the *it*+subclause construction. For all periods, the mean relative weight value for the preposed clausal subject construction is lower than the value for the *it*+subclause construction. For all periods there is also a slight difference when it comes to standard deviation (SD), where the variance of the relative weight values for the *it*+subclause construction is larger than that for the preposed clausal subject construction. As shown by a Wilcoxon rank sum test, the differences between the two constructions are statistically significant for all three periods.

Even though the data presented in Table 7.4 show clear differences between the two constructions, it cannot be concluded that the choice of construction can be fully predicted by these relative weight values. For the periods where the preposed clausal subject construction is relatively frequent (i.e. in Early and Late Modern English), there are several constructions with a very heavy subclause, but where this subclause nonetheless occurs as a preposed clausal subject. One of the most extreme examples is given in (5).

(5) [that thou should'st lie out by the way two Nights, and upon the Sunday get home, and there meet with this same black-bearded little Gentleman, and appoint these People to come to thy House upon the Tuesday; and when they came, entertain them three or four Hours at thy own House, and go back again so many Miles with

> them, and have no Entertainment but a piece of Cake and Cheese that thou broughtest thyself from home, and have no Reward, nor so much as know any of the Persons thou didst all this for], is very strange. (LISLE-E3-P1,4,112.556)

As we can see, the subclause here contains no less than 95 words, while the predicate consists of only three words. Nonetheless, the subclause participates in a preposed clausal subject construction. If relative weight was the sole factor determining the choice of construction, this sentence would clearly not have been expected.

7.3 Information structure

Turning next to matters of information structure, we may start by noting that the analysis of the relevant constructions derives from the theory of information structure as presented in Birner & Ward (1998), with an LFG-formalisation based on Dalrymple & Nikolaeva (2011). As outlined in Section 2.5, I assume the information structural notions given in Figure 7.2.

$$\begin{bmatrix} \text{STAT} & \{\text{IDENTIFIABLE, UNIDENTIFIABLE}\} \\ \text{ACTV} & \{\text{ACTIVE, INACTIVE, ACCESSIBLE, ANCHORED}\} \\ \text{GIVENNESS} & \{\text{GIVEN, NEW}\} \\ \text{CONTRAST} & \{\text{CONTRASTIVE, NONCONTRASTIVE}\} \end{bmatrix}$$

Figure 7.2: Attributes and values at i-structure

In the quantitative analysis of the corpus material, the notions of GIVENNESS and CONTRAST are investigated in different measurements.

7.3.1 Operationalisation of givenness

Recall from Section 2.5 that the notion of GIVENNESS concerns whether some proposition is assumed to be part of the information state of the addressee prior to the utterance, in which case it is GIVEN, or whether it assumed to be added to the information state after the utterance, in which case it is NEW. As it would have been too time-consuming to analyse and code each instance of the constructions in the corpora for this notion of givenness, a simpler operationalisation of givenness is required. For the purpose of this investigation, the influence of givenness on the

choice of construction is tested by considering whether the subclause in the constructions contains discourse-old material or not. If the subclause doesn't contain any discourse-old elements, it is coded as NEW. If it does, it is coded as UNDECIDED. The presence of discourse-old elements does not mean that the proposition expressed by the subclause as a whole is GIVEN. The label NEW does likewise not prefectly match the theoretical notion of NEW presented in Section 2.5. However, it is nonetheless reasonable to assume, disregarding Prince's category discourse-new hearer-old (Prince, 1988), that NEW subclauses, which do not contain any discourse-old elements, also are relationally new.

Three types of elements are recognised as discourse-old for the purpose of this investigation, whose presence thus indicates the category UNDECIDED. These are (i) pronouns and anaphoric adjectives or adverbs, (ii) definite NPs referring to a previously mentioned referent and (iii) verbatim repetitions of words. Examples of subclauses coded as UNDECIDED can be found in (6), taken from the BNC sample, where the discourse-old elements are marked in italics. Also, all three examples contain preposed clausal subjects.

(6) a. [Whether *such* a ferret fatality can be attributed to *it* remaining unmated] is highly debatable. [BNC]
b. [Whether disclosure is required if *the Calcutta office* bills *the US parent* direct] will depend on who requisitions the work. [BNC]
c. Laura Davies won the US Women's Open in 1987 and such is the power that she generates that any time she plays really well she wins. But [whether *she plays well*] appears to be a matter of chance. [BNC]

In (6-a), we have the personal pronoun *it* as well as the adjective *such*, both referring back to things mentioned in the previous discourse. In (6-b), there are two definite NPs, *the Calcutta office* and *the US parent*, the referents of which are both present in the previous discourse. In (6-c), finally, every word of the *whether*-clause is a verbatim repetition of something in the previous discourse. Accordingly, the subclauses in these three sentences are coded as UNDECIDED in the sample.

An example of a subclause without any discourse-old elements, which represents the *it*+subclause construction, is given in (7).

(7) Frequently one senses a certain unwillingness on the part of professional educators to accept these decisions and work with them. There is a fear that 'standards' may suffer, and a certain civil

> servant's distrust for 'politicians', bred of British traditions. It may also be asked [whether national political and economic policies are always debated and discussed with the seriousness they should be in curriculum committees or teachers' colleges]. [BNC]

Here, the *whether*-clause does not contain any elements mentioned or evoked in the previous discourse. It is thus coded as NEW. However, it is also relationally discourse-new in the sense that there is no proposition assumed to be present in the information state of the adressee in this particular sentence.

The coding of GIVENNESS was performed manually for the BNC sample, i.e. each sentence was coded for givenness according to the criteria described above. For the historical corpora, a coding query[2] was used in order to code givenness in the material. The presence of pronouns in the subclause was coded by searching for subclauses dominating an NP, which, in its turn, immediately dominates a pronoun (PRO). Non-indefinite noun phrases were found by searching for similar subclauses, but where an NP immediately dominates a determiner (D), which does not immediately dominate an element starting with the letter *a*, such as *a*, *an*. Included among the target determiners are *the*, *that*, *this*, *those*, *these*, *yon* and *yonder*.

7.3.2 Operationalisation of contrast

With respect to the phenomenon of contrast, only manual coding was possible. Such coding was thus only made for the BNC sample, which contains a manageable number of sentences for manual coding. The sentences in the sample were coded for the presence of a contrastive relationship between an element in the subclause and some other contextually available element. Consider the sentence in (8), which illustrates a case where there is a clear constrastive relationship between the proposition expressed in

[2]The coding query used for coding givenness is the one presented below. The label *pro* is used for pronouns, the label *det* for non-indefinite noun phrases and the label *such* for clauses containing the word *such*.

(i) //Givenness status of subclause
3: {pro: (IP-INF*|CP-THT*|CP-QUE* Doms NP*) AND (NP* iDoms PRO*)
det: (IP-INF*|CP-THT*|CP-QUE* Doms NP*) AND (NP* iDoms D*) AND (D* iDoms !a*)
such: (IP-INF*|CP-THT*|CP-QUE* Doms NP*) AND (NP* iDoms SUCH)
z: ELSE }

the preposed clausal subject and a proposition available in the previous discourse.

(8) I mean I don't write for therapeutic purposes in the sense that you might imagine, you know, someone in a mental hospital would paint or do pottery or conceivably write in order to relieve the inner tensions. ... But [whether it's therapeutic for the reader] is not for me to say. [BNC]

Evidently, there is a contrastive relationship here between the question whether *the writer writes* for therepeutic purposes and the question in the *whether*-clause on whether *the reader reads* the book for therepeutic purposes.

Relevant to note in this context is that a contrastive relation within the subordinate clause is not included in the coding. With respect to so called alternative questions, two alternative, and thus contrastive, questions are given in the subclause. Sentences with such subclause-internal contrast are not coded as contrast in the present investigation, since it does not concern the relation between the subclause and the context. Consider the sentence in (9).

(9) It is up to them [whether they move in with Mum and Dad or set up a caravan on site until the work is finished].

In (9), there is a contrastive relation within the *whether*-clause between the act of moving in with Mum and Dad, and the act of setting up a caravan on site.

Apart from coding for contrast or no contrast, there is a further distinction to be made here: (i) alternative contrast, concerning alternative propositions that differ in some regard, and (ii) polar contrast, concerning the truth or falsity of a proposition. Compare now the sentence in (8) above with the sentence in (10).

(10) Laura Davies won the US Women's Open in 1987 and such is the power that she generates that any time she plays really well she wins. But [whether she plays well] appears to be a matter of chance. [BNC]

In (8), there is a contrastive relation between the two alternative propositions, i.e. 'that an authors writes for therapeutic purposes' and 'that a reader reads for therapeutic purposes'. In (10), the proposition 'that she plays well' contrasts with the proposition 'that she doesn't play well'. In this particular example, the whole subclause, except the subordinator,

is repeated verbatim from the previous discourse, with a phonological emphasis on the word *whether*[3]. This gives rise to polar contrast. As will be seen from the results, the presence of polar contrast seems to be relevant in the choice of construction.

Having presented the operationalisation of the two phenomena givenness and contrast, then let us now turn to the concrete results of my investigation, starting with givenness and then proceeding to contrast.

7.3.3 Givenness in Present-Day English

Table 7.5 shows the quantitative results from the investigation of givenness with respect to the BNC sample.

Table 7.5: Frequencies of givenness in relation to the choice of construction

	UNDECIDED	NEW	Total
Preposed clausal subject	48	0	48
it+subclause	49	11	60
Total	97	11	108

(Fisher's Exact Test, p-value < 0.05, odds ratio = 2.28)

As can be seen in the table, the sample contains 97 *whether*-clauses coded as UNDECIDED and 11 coded as NEW. As for the cases coded as NEW, all of them feature the *it*+subclause construction. As for those coded as UNDECIDED, about half of these exhibit the preposed clausal subject construction and half of them the *it*+subclause construction. A Fisher exact test tells us that there is a significant difference with a small effect size between the two constructions in relation to the use of givenness.

An important observation to be made here is the total absence of instances of the preposed clausal subject construction coded as NEW. All instances coded as NEW represent the *it*+subclause construction. This fact supports the hypothesis put forward by both Miller (2001) and Birner & Ward (2004) that the preposed clausal subject is required to be given in some sense.

7.3.4 Givenness in the historical corpora

The data on givenness in the historical periods are based on the type of searches described in Section 7.3.1. As mentioned, when the subclause

[3]The emphasis on the word *whether* seems obligatory in the context of the sentence in (10).

contains a pronoun, definite determiner or the word *such*, the sentence is coded as UNDECIDED; otherwise it is coded as NEW. The results are given in Table 7.6.

Table 7.6: Sentences coded as UNDECIDED and NEW in the historical corpora

Middle English	UNDECIDED	NEW	total
Preposed clausal subject	32	27	59
it+subclause	980	412	1392
total	1012	439	1451

(Fisher's exact test, p-value < 0.05, odds ratio = 2.01)

Early Modern English	UNDECIDED	NEW	total
Preposed clausal subject	125	51	176
it+subclause	2434	900	3334
total	2559	951	3510

(Fisher's exact test, p-value = 0.60, odds ratio = NA)

Late Modern English	UNDECIDED	NEW	total
Preposed clausal subject	163	40	203
it+subclause	1479	672	2151
total	1642	712	2354

(Fisher's exact test, p-value < 0.05, odds ratio = 0.54)

The table shows some interesting results. First and foremost, there seems to be a diachronic shift in the data with respect to the occurrence of the preposed clausal subject construction. In Middle English, the proportion of sentences containing this construction coded as NEW and UNDECIDED is about the same, 50-50. In Early Modern English, the proportion is about 70 % coded as UNDECIDED and 30 % coded as NEW. In Late Modern English, the proportion is about 80 % coded as UNDECIDED and 20 % as NEW. In contrast, we may note that there were no instances of the preposed clausal subject construction coded as NEW in the BNC sample. Based on these data, it thus seems as if the preposed clausal subject construction is subject to some kind of functional specification, where sentences coded as NEW show a gradual decrease over time.

Further, as for the differences noted between the constructions in Table 7.6, for the Middle English period and the Late Modern English period, there is a statistically significant difference between the preposed clausal subject construction and the *it*+subclause construction when it

comes to the presence of pronouns, definite determiners and the word *such*. For these periods, there is a smaller proportion of sentences coded as *new* for the preposed clausal subject construction in comparison to the *it*+subclause construction. For the Early Modern English period, there is however no significant difference.

7.3.5 Contrast in Present-Day English

Flipping the coin over to contrast, we may start by considering the data in Table 7.7, which shows the frequencies of contrastive and noncontrastive *whether*-clauses in the BNC sample.

Table 7.7: Frequencies of contrast in relation to extraposition

	Contrastive	Noncontrastive	Total
Preposed clausal subject	15	33	48
it+subclause	2	58	60
Total	17	91	108

(Fisher's Exact Test, p-value < 0.05, odds ratio = 12.89)

There are 17 contrastive and 91 noncontrastive instances in the sample. Among the 17 subclauses coded as contrastive, 15 shows the preposed clausal subject construction and two the *it*+subclause construction. Among the 91 subclauses coded as noncontrastive, the distribution is more equal, with 33 featuring the preposed clausal subject construction and 58 the *it*+subclause construction. A Fisher exact test shows that there is a significant difference, with a strong effect, between the constructions in relation to contrastive and noncontrastive subclauses. Thus, from Table 7.7, we can conclude that contrast seems to have a significant effect on the choice of construction. Even more importantly, in terms of effect size for the difference between the constructions, the size of the effect is greater for contrast in comparison to givenness.

As hinted at in section 7.3.2, a further distinction to be made for the instances coded as contrastive is that of polar and alternative contrast. In the following table, the contrastive subclauses are divided up into subclauses expressing polar contrast and those expressing alternative contrast.

Table 7.8: Type of contrast in relation to the choice of construction

	Polar contrast	Alternative contrast	Total
Preposed clausal subject	8	7	15
it+subclause	0	2	2
Total	8	9	17

(Fisher's Exact Test, p-value = 0.47, odds ratio = NA)

Table 7.8 shows that all eight instances coded as expressing polar contrast relate to the preposed clausal subject construction. For instances coded as alternative contrast, both constructions are represented. The difference between the preposed clausal subject construction and the *it*+subclause construction with respect to type of contrast is not statistically significant, which means that it is difficult to draw any conclusions from this difference. It is nonetheless interesting to note that all instances of polar contrast represent the preposed clausal subject construction.

In the next section, a discussion is maintained of the relation between complexity, weight and information structure with respect to the data provided so far, including also particular examples.

7.4 Weight, complexity and information structure

While the two previous sections have considered weight, complexity and information structure in isolation, the present section will focus on these notions in relation to one another. The first subsection concerns the statistical correlations between relative weight, givenness and contrast. The second subsection maintains a discussion of individual sentences that seem to go against the tendencies observed in the quantitative data discussed earlier in the chapter.

7.4.1 Correlations between weight and information structure

Before we go into the discussion of particular examples, let us briefly consider the correlations between relative weight, givenness and contrast in the corpus data. As has been seen throughout this chapter, each variable, with the exception of givenness in the Early Modern English period, seems to have a significant effect on the choice between the preposed clausal subject construction and the *it*+subclause construction. In this

section, the central question is raised whether each variable still might have a significant effect on the choice of construction when controlling for the other variables. To this end, multiple regression statistics (Gries, 2009: 291-306) will be used, where relative weight, contrast and givenness form predictor variables and where the dependent variable is the choice between the preposed clausal subject construction and the *it*+subclause construction.

Let us start with the data from the BNC sample. Without going into too much detail, using a binary logistic regression test, with a statistical model[4] containing the variables relative weight, givenness and contrast, two things are found: (i) there is a significant correlation, controlling for the other variables, between the choice of construction and the two variables relative weight and contrast, (ii) there is no significant correlation in the case of givenness. This means among other things that the higher the relative weight value, the more the *it*+subclause construction is preferred (odds ratio≈16.88, 95 %-CI 3.37 and 115.15, p-value<0.05). If instances are coded as noncontrastive, there is also a preference for the *it*+subclause construction (odds ratio≈8.72, 95 %-CI 2.05 and 60.65, p-value<0.05). For givenness, on the other hand, there is only a nonsignificant preference for the *it*+subclause construction, when instances are coded as new (odds ratio≈6.36, 95 %-CI 1.04 and 125.74, p-value<0.05).

For the historical data, similar regression tests for the three periods Middle English[5], Early Modern English[6] and Late Modern English[7] show, on the one hand, that relative weight has a significant effect on the choice of construction for all three periods, when the factor of givenness is controlled for; on the other hand, givenness only has a significant effect in the Late Modern English period, when the factor of relative weight is controlled for. The statistical models for the historical periods account for very little of the variation of the data. While the model for Present-day English accounts for about 38 % of the variance (R^2=0.378), which is in itself not very much, the models for the historical data account for between 6 % and 8 %, which is very little. Part of the difference could be due to the measurements for weight and givenness being less refined for the historical material, and the factor of contrast not being included.

Further, for Middle English, a higher relative weight correlates with more *it*+subclause constructions (odds ratio≈29.53, 95 %-CI 8.06 and 113.11, p-value<0.05). If instances are coded as new, there is a non-

[4]Log-likelihood ratio χ^2=35.78, df=3, p<0.05, R^2=0.378, C=0.783, D_{xy}=0.565.
[5]Log-likelihood ratio χ^2=31.76, df=2, p<0.05, R^2=0.081, C=0.714, D_{xy}=0.427.
[6]Log-likelihood ratio χ^2=70.02, df=2, p<0.05, R^2=0.070, C=0.699, D_{xy}=0.397.
[7]Log-likelihood ratio χ^2=59.08, df=2, p<0.05, R^2=0.063, C=0.674, D_{xy}=0.349.

significant decrease in the probability of the *it*+subclause construction being used (odds ratio≈0.62, 95 %-CI 0.35 and 1.08, p-value=0.09). The fact that there is a *decrease* is surprising, going against as it does what the Present-day English data show. However, as the trend is non-significant, no safe conclusions can be drawn.

For Early Modern English, a high relative weight value similarly increases the probability for the *it*+subclause construction to occur (odds ratio≈18.83, 95 %-CI 9.27 and 39.05, p-value<0.05), while givenness shows only non-significant effects (odds ratio≈0.92, 95 %-CI 0.65 and 1.32, p-value=0.63).

In Late Modern English, finally, the effect of relative weight is similar to that of the previously described periods, although not that strong (odds ratio≈9.08, 95 %-CI 4.69 and 17.71, p-value<0.05). In contrast to the other periods of historical English, however, instances coded as new in the Late Modern English period significantly increases the preference for the *it*+subclause construction (odds ratio≈2.20, 95 %-CI 1.52 and 3.26, p-value<0.05).

Thus, having looked at the various statistical patterns, we can conclude that, although each variable has a significant effect on the choice of construction on its own, there are differences between the variables in the extent to which they exhibit a significant correlation. Relative weight exhibits a significant correlation in all periods. For the Present-day English material, both relative weight and contrast exhibit a significant correlation, while givenness does not. As for the specific behaviour of givenness, it does not exhibit a significant effect for Middle English and Early Modern English, but it does for Late Modern English.

These results are likely to be partly a consequence of the way in which the different variables have been operationalised. As discussed in Section 7.3.1, there is a far from perfect match between the theoretical notion of givenness and the way in which this notion has been operationalised. However, while the variables relative weight and contrast in the Present-day English material have been dealt with in a relatively fine-tuned way, the variable givenness correponds only in a very rough way to the theoretical notion of givenness. This is to say that the fact that givenness here constitutes only a rough estimate could have had a certain effect on the results of the statistical investigation. Thus, as a consequence of the way the factors have been operationalised, a more detailed investigation of individual examples is probably necessary in order to find the true relationship between relative weight and information structure in the material. Accordingly, this is what is attempted in the next two sections.

7.4.2 Weight, complexity and information structure in PDE

After having considered the statistical relationship between weight/complexity and givenness/contrast and their influence on the choice of construction, let us now look at some of the instances that seem to go against the generalisations about weight, complexity, givenness and contrast presented in the two previous sections. In such cases, we would expect there to be a competition between some of the factors in the individual example. Based on the consideration of the sentences that go against expectations in the BNC sample, we will see that the following decision tree emerges.

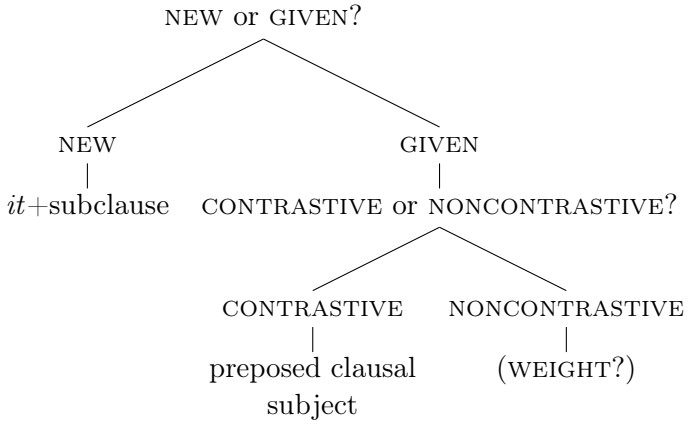

Figure 7.3: Decision tree on the influence of information structure and weight on the choice of construction

In a situation where there is a clash, it seems as if givenness (NEW or GIVEN) is most decisive in the choice of construction, followed by contrast (CONTRASTIVE or NONCONTRASTIVE) and WEIGHT.

As we have seen, it seems as if relationally NEW subclauses exclusively feature the *it*+subclause construction. They do not occur as preposed clausal subjects, making it a necessary condition on the subclause in the preposed clausal subject construction that it is relationally GIVEN. However, relationally GIVEN subclauses seem to occur in both constructions (cf. Kaltenböck, 2005). Examples of relationally GIVEN subclauses in the two constructions are given in (11).

(11) a. Look, Guido will never know whether we're studying or not. He's so preoccupied with his wretched speedboat race... Every afternoon he's out at sea doing practice runs... That's

all he ever thinks about these days.' Little minx, Ronni thought, as she looked into Silvia's cajoling face. So Guido hadn't been entirely wrong, after all! She said, 'It doesn't matter [whether Guido knows or not]. I'm contracted to tutor you for so many hours a day, and I'm afraid I must insist on fulfilling my contract.' [BNC]

b. Laura Davies won the US Women's Open in 1987 and such is the power that she generates that any time she plays really well she wins. But [whether she plays well] appears to be a matter of chance. [BNC]

In (11-a), a relationally GIVEN subclause features the *it*+subclause construction. The question of whether or not Guido knows that they are studying is explicitly mentioned in the previous discourse. In (11-b), a relationally GIVEN subclause instead occurs as a preposed clausal subject. The question whether 'she plays well' is a verbatim repetition of part of the previous discourse.

The fact that relationally GIVEN subclauses feature both constructions, while NEW subclauses only feature the *it*+subclause construction, makes the status of the subclause as GIVEN a necessary, but not sufficient condition in the choice of the preposed clausal subject construction. We will see here that contrast also seems to play an important part. In the discussion below, support from more individual sentences will be given for the decision tree in 7.3.

In the decision tree, the second question concerns whether the subclause is CONTRASTIVE or NONCONTRASTIVE. Subclauses that are CONTRASTIVE are assumed to be realized as preposed clausal subjects. As can be recalled from Table 7.7, 15 out of the 17 cases coded as CONTRASTIVE were preposed clausal subject constructions. However, there were also two cases of contrast noted in connection with the *it*+subclause construction, which thus would not be expected. As will be seen, the subclauses in these two instances, given in (12), are, on a closer inspection, to be analysed as NEW, and thus precluded from being realized as preposed clausal subjects.

(12) a. What Federico Cesi once wrote of the members of his Lincean Academy applied to Galileo with a vengeance: 'All we claim in common is freedom to philosophize in physical matters'. There can be no denying this was an issue. It is doubtful [whether it was ever such an issue for Wilkins]. Like Galileo, he was committed to the Copernican system as a cosmology and not merely as a mathematical hypothesis. [BNC]

b. It seems that, in evaluating the significance of pre-S protein display in liver in chronic hepatitis B virus infection, special emphasis should be made on their topographical (cytoplasmic or membranous) distribution as well as the quantitative expression, as suggested in the study of intrahepatic expression of HBsAg and HBcAg in chronic type B hepatitis. Furthermore, it has long been suggested that there is an inhibitory effect on hepatitis B virus replication by hepatitis delta virus in chronic hepatitis B virus infection, but it remains unclear [whether hepatitis delta virus might interfere with the expression of hepatitis B virus envelope antigens in the liver or not]. [BNC]

In (12-a), there is a contrastive relation between something being an issue for Wilkins and this something being an issue for someone else. In (12-b), there is a corresponding contrastive relation between the inhibitory effect on virus replication of hepatitis delta virus and that on the expression of antigens by hepatitis delta virus. While all other instances coded as CONTRASTIVE correlate with the preposed clausal subject construction, these two cases favour the *it*+subclause construction. In terms of relative weight, both instances have values favouring extraposition. The example in (12-a) has a relative weight value of 0.6, while the example in (12-b) has a relative weight value of 0.8. Assuming that CONTRAST is more influential than WEIGHT in the choice of construction, we would, however, not expect instances coded as CONTRASTIVE in the *it*+subclause construction, unless there is some other factor that is even more important. As is represented in the decision tree, there seems to be such a factor, namely GIVENNESS. Although the phrase *for Wilkins* is interpreted as CONTRASTIVE, on a closer inspection, the proposition that 'it was ever such an issue for Wilkins' seems to be relationally NEW. The referent of *Wilkins* is not present is the immediately preceding discourse, and the question of whether something is an issue for him does not seem to have been discussed previously. In (12-b), the subclause also appears to express a relationally NEW proposition, despite the presence of CONTRAST. The question whether hepatitis delta virus interferes with the expression of this type of antigens is introduced as something which is unclear, and which hence requires further investigation. Thus, both instances contain what seems to be relationally NEW subclauses. Since subclauses whose content is relationally NEW do not go together with preposed clausal subjects, the *it*+subclause construction is the only option.

With respect to CONTRAST, another observation needs to be discussed, supporting the decision tree suggested in Figure 7.3. When the subclause expresses polar contrast between the truth and falsity of a proposition forming part of the context, we have already seen that all examples favour the preposed clausal subject construction, despite their relative weight. The reason for this seems to be connected to the relationship between polar constrast and givenness. Consider now the sentence in (13).

(13) At their meeting in either 1749 or 1750, Christopher Smart accepted this task from Richardson; however, he delayed writing the piece which was to be printed in the volume. Indeed with most of the book already printed, Richardson had not by December 10,1750 received the epitaph. [Whether the epitaph that was printed is Smart's work at all] is unknown. [BNC]

While this sentence has a high relative weight value (0.69), it nonetheless selects the preposed clausal subject construction. It seems here as if the question of whether 'the epitaph that was printed is Smart's work at all' is taken to be part of the information state of the addressee. The subclause expresses a polar contrast between the truth or falsity of this proposition. The presence of polar contrast seems to entail that the proposition expressed is GIVEN. Accordingly, the instances coded as polar contrast are both GIVEN and CONTRASTIVE, a combination that seems especially prone to trigger the preposed clausal subject construction.

After weight and givenness, we finally return to the influence of weight on the choice of construction. As could be seen in Figure 7.1 in Section 7.2.2., for the lower relative weight values, between -0.8 and +0.2, the great majority of instances are preposed clausal subject constructions. Out of 20 instances, 17 contain preposed clausal subjects and only three the *it*+subclause structure. In (14), these three instances of the *it*+subclause construction are given in conjunction with part of the preceding discourse.

(14) a. What clearly are wanted are the links that people cherish with their traditional counties. Even if the units of government are not based upon the counties, people still hark back to the links that they had with the traditional counties. They are still keen on those links. There is no reason why the traditional counties should not emerge as part of the process, even if they are counties that have no administrative functions in certain places. The Opposition laugh at their peril. It matters very much indeed to the people of Wirral [whether they live in Cheshire or Merseyside]. It matters very much

to the people of Coventry whether they live in Warwickshire or in the west midlands. [BNC]

b. It is an interesting thought. Whether Henry Eliot's letter is still on some Home Office file or other among the seventeen miles of paper which Toynbee (if I remember right) calculated to comprise the total government documentation of the war, I do not know; but some research worker may conceivably yet come upon it, and it would be interesting to learn [whether it contained any apt comments]. [BNC]

c. This resulted in Blacks overall receiving proportionately more custodial sentences. It appeared, on examining the offences involved, that Blacks had more' indictable only' offences, and also more' (——) for which they were committed for trial. It was not possible to know from the data available [whether this was the defendants' choice or magistrates declining to try them]. [BNC]

The sentences in (14) are *it*+subclause constructions, despite the fact that their weight values are statistically linked to the preposed clausal subject construction. For these sentences, it seems as if it is givenness that determines the choice of construction. In (14-a), the choice of the *it*+subclause construction is supported by the fact that the content of the subclause is NEW, and that there are no discourse-old elements in it. The sentences in (14-b) and (14-c), on the other hand, both contain some discourse-old elements, and are thus coded as UNDECIDED. In (14-b), the pronoun *it* refers back to the documentation of the war, and, in (14-c), the demonstrative pronoun *this* refers back to the proposition that 'Blacks had more "indictable only" offences'. However, despite the fact that they contain discourse-old elements, on a more fine-grained analysis, the subclauses in these sentences both seem to be relationally NEW. In (14-b), the proposition that 'the documentation contains any apt comments' is not presupposed to be given information. The same holds for the sentence in (14-c), where the proposition about possible reasons for *this* is not presupposed to be given information. Thus, despite the fact that the weight values favour preposed clausal subjects, closer inspection reveals that the content of all three subclauses in (14) must be analysed as NEW, which seems to be disallowed in the preposed clausal subject construction.

In conclusion, the data support the decision tree in 7.3, where GIVENNESS is most infuential in the choice of construction, followed by CONTRAST. Relationally NEW subclauses occur exclusively in the *it*+subclause construction. Subclauses expressing a contrastive relation select the pre-

posed clausal subject construction, unless they are relationally NEW. Finally, when both constructions are possible, WEIGHT seems to be an important factor for the constructional choice.

7.4.3 Weight, complexity and information structure in historical English

Following the discussion of individual sentences from the BNC sample in the preceding section, this section discusses individual sentences from the historical corpora concerning the relationship between weight, complexity and information structure.

In the historical data, there are several instances where the choice of construction differs from the expectations based on generalisations about relative weight, i.e. that low relative weight, particularly values below +0.3, goes together with preposed clausal subjects, while high relative weight, above +0.3, is associated with the *it*+subclause construction. In the present section, we will consider two thought-provoking examples: one is a preposed clausal subject construction with a high relative weight value, and the other is an *it*+subclause construction with a low relative weight value. These two instances will be used to illustrate the relationship between relative weight and information structure in the choice of construction in the historical material. Both examples derive from the Early Modern English period. As not all the sentences in the historical data have been considered individually, the discussion therefore only serves to give a glimpse into the imformation structure of the historical data. Further research is required to determine whether the tendencies observed hold more generally.

The first example is the sentence previously given in (5), here repeated as (15), now provided together with some additional context. As will appear, this example has an extremely high relative weight value, +0.94, but is nonetheless a preposed clausal subject construction. The example consists of a dialogue between two individuals, L.C.J and Dunne. Dunne has just finished answering questions about how he carried a message to a woman.

(15) L. C. J. And this is as much as you know of the Business?
Dunne. Yes, my Lord, this is all that I remember.
L. C. J. Well; and what hadst thou for all thy pains?
Dunne. Nothing but a Month's Imprisonment, my Lord.
L. C. J. Thou seemest to be a Man of a great deal of Kindness and Good-nature; for, by this Story, there was a Man that thou

> never sawest before for I would fain have all People observe what Leather some Men's Consciences are made of and because he only had a black Beard, and came to thy House, that black Beard of his should persuade thee to go 26 Miles, and give a Man half a Crown out of thy Pocket to shew thee thy way, and all to carry a Message from a Man thou never knewest in thy Life, to a Woman whom thou never sawest in thy Life neither; [that thou should'st lie out by the way two Nights, and upon the Sunday get home, and there meet with this same black-bearded little Gentleman, and appoint these People to come to thy House upon the Tuesday; and when they came, entertain them three or four Hours at thy own House, and go back again so many Miles with them, and have no Entertainment but a piece of Cake and Cheese that thou broughtest thyself from home, and have no Reward, nor so much as know any of the Persons thou didst all this for], is very strange. (LISLE-E3-P1,4,112.549-556)

The content of the preposed clausal subject in (15) is clearly GIVEN. It is a recollection of events that the addressee himself has told the speaker. The use of a preposed clausal subject is thus in line with Ward & Birner's (2004) constraint, i.e. that the content of the preposed clausal subject must be at least hearer-old. However, as discussed in the immediately preceding section, the fact that the content of the subclause is GIVEN does not explain the choice of construction, since relationally GIVEN subclauses can occur in either construction. Arguably, the choice of the preposed clausal subject construction here indicates an interpretation where the content of the subclause is taken to express polar contrast. After all, it is possible to rephrase the example as: 'To do all this, rather than not doing it, is very strange'. Thus, despite an extremely high relative weight value, the presence of a contrastive relation nonetheless seems to lead to the choice of the preposed clausal subject construction.

As an offset to this, consider now the *it*+subclause construction in (16). Here, it instead seems to be the lack of a contrastive relation that determines the choice of construction.

(16) And ye wicked goaler: one Hunter: a younge man: hee would come & give ye horse a whippe & make him skippe & leape: & then hee would come & looke mee in ye face & say: how doe you M=r= ffox: but I tolde him it was not civill In him [to doe soe]. (FOX-E3-P1,93.91)

This instance has a relative weight value of -0.25, which is a value favouring the preposed clausal subject construction. Furthermore, the content of the subclause represents GIVEN information, as *to do so* refers back to the act of 'looking him in the face and saying "how do you do, mr Fox"'. As pointed out above, relationally GIVEN subclauses may occur in either construction, but still we would not expect an instance with this relative weight value to occur in the *it*+subclause construction. The subclause does not express a contrastive relation, which would have been something that could have triggered the preposed clausal subject construction. Thus, in this sentence, none of the factors of givenness, contrast or relative weight seem to determine the choice of construction. However, just as for Bolinger's example in (12), discussed in Chapter 6, it here seems as if it is the lack of a contrastive relation that supports the choice of the *it*+subclause construction. Tentatively, it could thus be suggested that the combination between GIVEN and NONCONTRASTIVE leads to the choice of the *it*+subclause construction.

Not all the sentences from the historical corpora have been considered individually, which makes it hard to draw any general conclusions. However, as was illustrated in the discussion above, it seems as if similar constraints on givenness and contrast are at work in the historical data, just as they are in Present-day English. Nonetheless, further research is required to ascertain to what extent this conclusion holds for a larger material. Based on the quantitative data presented in 7.3.4., it seems as if the influence of GIVENNESS becomes more and more pronounced over time.

7.5 Summary

To sum up, the present chapter has dealt with data on weight, complexity and information structure in relation to the choice between preposed clausal subjects and the *it*+subclause construction. Two kinds of material have been used. With respect to Present-day English, the BNC sample of 108 *whether*-clauses participating in the relevant constructions has been used. With respect to historical English, *that*-clauses, *wh*-clauses

(including *whether*-clauses) and infinitival clauses formed part of the material.

When it comes to the choice of construction in the BNC sample, both weight/complexity and givenness/contrast seem to have a significant effect. In particular, the *it*+subclause construction shows significantly higher relative weight values. In terms of the IC-to-word ratio, 100 out of 108 constructions would have had amore favourable value as *it*+subclause constructions. As for information structure, all sentences coded as NEW favoured the *it*+subclause construction, thus testifying to the influence of givenness. As for contrast, 15 out of 17 sentences were realised as the preposed clausal subject construction. On the whole, it seems as if relationally NEW subclauses occur exclusively in the *it*+subclause construction, and that subclauses expressing polar contrast occur exclusively in the preposed clausal subject construction. On the basis of these observations, it is possible to posit a decision tree, where the realisation of the subclause as GIVEN or NEW is the most influential in the choice of construction, followed by the realisation of the subclause as CONTRASTIVE or NONCONTRASTIVE. When information structure does not determine the choice if construction, weight seems to have a considerable effect.

With respect to the choice of construction in the historical data, similar factors seem to be at play. Thus, in all historical periods, the *it*+subclause constructions show significantly higher relative weight values. The influence of givenness on the choice of construction seems to increase over time. In Middle English, about 50 % of the preposed clausal subject constructions are coded as GIVEN. In Late Modern English, the percentage is even higher, about 80 %.

Part IV

Conclusions and future research

Chapter 8

Conclusions and future research

This dissertation has dealt with different aspects of the alternation between the preposed clausal subject construction and the *it*+subclause construction. Chapters 4 and 5 concerned the syntax and argument structure of these constructions, and Chapters 6 and 7 concerned the influence of weight, complexity and information structure on the choice of construction. The present chapter gives a summary of the conclusions arrived at and some pointers for future research.

8.1 Conclusions

The main conclusions offered in this dissertation are the following:

- The syntactic and morphosyntactic subject properties of subclauses in the history of English:
  - Subclauses in Old English do not occur as preposed clausal subjects, except in sentences constituting rigid translations from Latin. Preposed clausal subjects in non-Latin-based texts are first attested during the Middle English period.
  - Clause-final subclauses act as morphosyntactic subjects from the Old English period onwards, with decreasing frequency.
  - All preposed subclauses in Early and Late Modern English act as morphosyntactic subjects, but only infinitival clauses can occasionally be analysed as being structural (syntactic) subjects.

- Properties of a propositional subclause in conjunction with a subject *it*:
  - Two types of *it*+subclause constructions are found from late Middle English onwards: (i) *it*+ADJ and (ii) *it*+COMP.
  - In Old English, there was no *it*+COMP construction.
- Pragmatic and processing-related aspects of the alternation between preposed clausal subjects and the *it*+subclause construction:
  - The instances of the *it*+subclause construction in the BNC sample generally have a more favorable IC-to-word ratio (complexity value) in comparison to the instances of the preposed clausal subject construction. In Middle English, Early Modern English and Late Modern English, the subclause in the *it*+subclause construction generally represents a larger proportion of the sentence in comparison to the preposed clausal subject construction.
  - The choice between the preposed clausal subject construction and the *it*+subclause construction in the BNC sample is partly the result of considerations of information structure, where discourse-new subclauses exclusively occur in the *it*+subclause construction and where subclauses showing polar contrast exclusively occur in the preposed clausal subject construction.
  - In Middle English, Early Modern English, and Late Modern English, whether the subclause contains discourse-old elements is increasingly (over time) relevant for the choice of construction.
  - For Present-Day English, a decision tree can be made in which the realisation of the subclause as GIVEN or NEW is what is most influential on the choice of construction, followed by the realisation of the subclause as CONTRASTIVE or NONCONTRASTIVE. Although the results are more tentative for the historical data, roughly the same tendencies can be seen.

In the rest of this chapter, the conclusions offered above are presented in more detail, followed by some pointers for future research.

8.1.1 Clausal subjects and extraposition

The structural and functional subject properties of subclauses have been discussed in the linguistic literature for a long time. The corpus data

discussed in Section 5.1 indicates that there are differences between functional (morphosyntactic) and structural (syntactic) subject properties of subclauses in the history of English. In Early and Late Modern English, it seems as if subclauses have a number of functional subject properties, but no corresponding structural properties, except in the case of infinitival clauses. With respect to functional properties, all types of subclauses in Early and Late Modern English are attested in subject raising and coordinate subject deletion. The evidence from verb agreement and control is inconclusive in this respect. There are no examples of control and the examples of verb agreement all contain a postverbal plural NP, which could be responsible for the plural morphology on the verb.

In terms of structural subject properties, the corpus data show that *that*-clauses and *wh*-clauses (including *whether*-clauses) do not occur in the position between a fronted phrase, or a subordinator, and the finite verb. When it comes to infinitival clauses, they occasionally do occur in these environments in Early and Late Modern English, attesting to the grammaticality of infinitival clauses as structural subjects in these periods. Whether the lack of subclauses in subject position is due to ungrammaticality or extragrammatical factors is hard to say. As was discussed in Chapter 6, the cases where a clausal subject unambiguously occurs in a subject position are also cases of center-embedding, a situation which seems to be dispreferred for reasons of processing.

For the Old and Middle English data, discussed in Section 5.3, there is very little evidence available for the determination of the grammaticality of subclauses as subjects. With respect to preposed clausal subjects, there are only four examples in the Old English prose corpus. Furthermore, as pointed out in Section 5.3.1., all four examples are translations from Latin following the Latin word order rigidly. With respect to clause-final subclauses, the question of whether they can be analysed as subjects remains unresolved. However, based on the parallel between NP arguments and clausal arguments, an analysis of the subclause as a subject in passive constructions, where there is no preverbal subject constituent, does provide a more economic account. Nonetheless, with the theory of argument structure presented here, it would also be possible to analyse such sentences as subjectless, where the argument expressed as a subclause is demoted to COMP. Which analysis is preferable is a matter for future research to determine.

One of the principal points made in the dissertation concerns the analysis of the *it*+subclause construction in Section 5.2. It is proposed that this construction can be divided into two: (i) *it*+ADJ and (ii) *it*+COMP.

It+ADJ has a thematic subject *it* and an adjunct subclause, while *it*+COMP has a non-thematic subject *it* and a complement subclause. This distinction is based on data from extraction and what different constructions the relevant predicates participate in. *It*+COMP occurs with raising predicates and with the copula *be* in passive constructions. What these predicates have in common is that the first argument, arg1[–r], is not associated with a thematic role. Apart for accounting for differences in the possibility of extraction out of the subclause between raising and non-raising predicates, the analysis proposed also provides an explanation for the phenomenon known as obligatory extraposition. The fact that a verb such as *seem* only takes a preposed clausal subject when there is a secondary predicate follows from the claim that this verb takes two argument slots, rather than one.

Diachronically, the claim is that the *it*+COMP construction emerges in conjunction with the development of raising verbs in the Middle English period, while the *it*+ADJ construction is available in all periods of the history of English. This claim is based on the analysis of the behaviour of various verbs in the history of English. In the Old English period, concerning the verbs *þyncan* and *gelimpan*, semantic counterparts to the Present-day English raising verbs *seem* and *happen*, there are two things that point to the analysis of the *it*+subclause construction during this time as *it*+ADJ. Firstly, a subject *it*, when it occurs with these verbs during the Old English period, consistently occurs in addition to the thematic constituents of the clause. When there is a non-thematic subject, we would expect a subject *it* to replace a thematic subject rather than occur in addition to it. With respect to the verb *þyncan*, a subject *it* simply occurs in addition to the subclause, pushing the subclause out of the list of arguments into the role of adjunct. Secondly, the *it*+subclause construction for these verbs does not seem to alternate with the raising construction. No true examples of raising are found with these predicates are found. The alternation between the *it*+subclause construction and the raising construction for this group of predicates seems to emerge during the Middle English period.

8.1.2 Weight, complexity and information structure

In Part III, Chapters 6 and 7, the syntactic analysis given in Part II is supplemented by a discussion of the weight, complexity and information structure of the alternation between preposed clausal subjects and the *it*+subclause construction.

As discussed in Section 7.2., there are significant correlations between weight/complexity values and the choice of construction both in the BNC sample and in the historical corpora. Both the IC-to-word values and the relative weight values for the preposed clausal subject construction in the BNC sample are considerably lower than those for the *it*+subclause construction. With respect to the IC-to-word values, in 100 out of 108 occurrences, the choice of the *it*+subclause construction would lead to a more favourable complexity value. With respect to the historical data, the preposed clausal subject construction similarly has significantly lower relative weight values in comparison to the *it*+subclause construction, which means that the subclause in the preposed clausal subject construction typically constitutes a smaller proportion of the sentence in comparison to the *it*+subclause construction.

The results of the investigation of the information structure of the preposed clausal subject construction in relation to the *it*+subclause construction are discussed in Section 7.3. For the BNC sample, both givenness and contrast have significant effects on the choice of construction. All subclauses coded as NEW occur in the *it*+subclause construction. In terms of contrast, 15 out of 17 subclauses coded as CONTRASTIVE occur in the preposed clausal subject construction. Based on a closer examination of the two constrastive examples that do not occur in the preposed clausal subject construction, it turns out that they are relationally NEW. Thus, it seems as if the idea, which is also expressed in Miller (2001) and Ward & Birner (2004), that preposed clausal subjects are required to be relationally GIVEN can be supported. However, the fact that the subclause is GIVEN is not sufficient to determine the choice of construction. A consideration of individual examples in Section 7.4.2 shows that presence of a contrastive relation for those subclauses that are not NEW seems to be required for the choice of the preposed clausal subject construction. Thus, it seems possible to make a decision tree in which the realisation of the subclause as GIVEN or NEW is what is most influential on the choice of construction, followed by the realisation of the subclause as CONTRASTIVE or NONCONTRASTIVE.

In historical English, similar considerations seem to be at play. Interestingly, the influence of givenness on the choice of construction in the historical corpora seems to increase over time. In Middle English, about 50 % of the preposed clausal subject constructions comtain discourse-old elements. In Late Modern English, the percentage is higher, about 80 %.

8.2 Future research

In this final section, I present some questions for future research, including two problems concerning passive constructions, the question of the influence of prosody, and the nature of the Subject Condition.

The first problem requiring further investigation concerns passive sentences such as the one in (1), repeated from Section 5.2.4.

(1) To these things must be added, that moral Obligations can extend no further than to natural Possibilities.
(BUTLER-1726,241.108)

The sentence in (1) constitutes a passive construction with a subclause occurring in clause-final position, which lacks a subject *it*. The question here is what makes up the subject: the subclause or the initial lexically governed phrase *to these things*. Considering the importance of structural position as an indication of subjecthood in Early and Late Modern English, we would not expect to find sentences such as the one in (1) during this period[1].

A second problem concerns the difference in argument structure between a verb such as *say* and verbs such as *expect* and *believe*. As pointed out in Section 5.2.1, the predicate *say* participate in the passive raising construction, but not in the active subject-to-object raising construction. Verbs such as *expect* and *believe* participate in both constructions. Furthermore, as shown in (2), the verb *say* does not participate in the active control construction, while the verb *expect* does.

(2) a. they never expected to see us any more during the voyage
(COOK-1776,35.730)
b. *they said to see us during the voyage
'They said that they saw us during the voyage.'
[constructed]

In the active sentence of (2-a), the verb *expect* occurs in an active control construction, where the thematic subject of *expect* controls the subject of the non-finite verb *see* within the infinitival clause. The verb *say*, on the other hand, does not occur in control constructions, as can be seen in (2-b). What explains the fact that the verb *say* does not take an XCOMP in the active while it does so in the passive?

The third area of future research, which is tangential to the discussion about weight and complexity in Chapter 7, concerns the influence of

[1] For an analysis of similar constructions in Present-day Swedish, see Engdahl (2012).

prosody on the alternation investigated. Light (2012: 164), for example, speculates that extraposition (relative clause extraposition) might be motivated by prosodic wellformedness. According to her theory, the extraposition of the relative clause allows the main clause and the relative clause to form separate intonation phrases, which, according to Light, is deemed preferable prosodically. Further research is required to determine whether this could also be relevant for the alternation under discussion here. It is possible that the tendency for the subclause to occur at the clause periphery, either in the preposed clausal subject construction or in the *it*+subclause construction, might be motivated by the preference for the subclause to form its own intonation phrase.

Furthermore, with respect to preposing, Light (2012) refers to Büring (2007), who claims that contrastive topics are typically realised as their own intonation phrases. According to Light (2012: 167), this motivates the movement of phrases marked as contrastive topics to the left periphery. In Chapter 7, it was posited that the preposing of the subclause gives rise to presuppositions about givenness and contrast. This raises the interesting question of whether the preposing of the subclause might be phonologically motivated rather than motivated directly by information structure.

Yet another area for future research concerns the nature of the Subject Condition. This question was briefly discussed in Chapter 2, sketching two formulations of the Subject Condition: 'every predicator must have a subject' (Bresnan et al., 2016: 334) and 'every verbal predicate must have a SUBJ' (Dalrymple, 2001: 19). Within Chomskyan syntax, the Subject Condition can be compared to the EPP condition (Chomsky, 1981). Similar to the Subject Condition, but formulated within a different framework, the EPP condition says that certain configurations must have subjects (Chomsky, 1981: 27). According to Lasnik (2003), one of the original motivations for the EPP condition comes from raising constructions. In raising verb alternations, such as in Lasnik's examples given in (3), the EPP condition is needed to explain the requirement for the subject position of the raising verb to be filled.

(3) a. It seems that John is here.
b. *Seems that John is here.

Assuming that the verb *seem* does not assign a subject theta role, the EPP condition explains the ungrammaticality of (3-b). In LFG, the Subject Condition has also been involved in explaining the ungrammaticality of sentences such as (3-b). Assuming that something precludes the *that*-

clause from being mapped to SUBJ, a non-thematic subject is needed to satisfy the subject condition. As was discussed in Chapter 5, a different analysis is here given for the sentences in (3), in which raising verbs generally take two syntactic arguments, a claim that does not involve the Subject Condition. As a result of Kibort's revised Lexical Mapping Theory, the Subject Condition is rendered redundant. Further research is needed, however, to estimate the full implications of a theory not recognising a Subject Condition.

In the present section, a few areas for future research have been identified. There are of course many more unresolved issues in the study of these constructions, both with respect to their syntax and argument structure and with respect to the factors involved in the choice of construction.

Bibliography

Alejchem, Scholem. 1922. *Aus dem nahen Osten*. Berlin/Wien: Benjamin Harz Verlag.

Allen, Cynthia. 1986. Dummy Subjects and the Verb-Second 'Target' in Old English. *English Studies* 67(6). 465–470.

Allen, Cynthia. 1995. *Case Marking and Reanalysis: Grammatical Relations from Old to Early Modern English*. Oxford: Oxford University Press.

Alrenga, Peter. 2005. A Sentential Subject Asymmetry in English and Its Implications for Complement Selection. *Syntax* 8(3). 175–207.

Alsina, Alex, Tara Mohanan & KP Mohanan. 2005. How to Get Rid of the COMP. In Miriam Butt & Tracy Holloway King (eds.), *The Proceedings of the LFG05 Conference*, Stanford, CA: CSLI Publications.

Anderson, John M. 1988. The Type of English Impersonals. In John M. Anderson & Norman Mcleod (eds.), *Edinburgh Studies in the English language*, 1–22. Edinburgh: John Donald.

Anderson, John M. 1997. Preliminaries to a History of Sentential Subjects in English. *Studia Anglica Posnaniensia XXXI* .

Arnold, Jennifer E., Anthony Losongco, Thomas Wasow & Ryan Ginstrom. 2000. Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering. *Language* 1. 28–55.

Asudeh, Ash. 2012. *The Logic of Pronominal Resumption*. Oxford: Oxford University Press.

Attia, Mohammed. 2008. A Unified Analysis of Copula Constructions in LFG. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG08 Conference*, Stanford: CSLI Publications.

Barron, Julia. 1997. LFG and the History of Raising Verbs. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG97 Conference*, Stanford, CA: CSLI Publications.

Barron, Julia. 2001. Perception and Raising Verbs: Synchronic and Diachronic Relationships. In Miriam Butt & Tracy Holloway King (eds.), *Time over matter: Diachronic perspectives on morphosyntax*, Stanford, CA: CSLI Publications.

Behaghel, Otto. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25. 110–142.

Berman, Judith. 2003. *Topics in the Clausal Syntax of German.* Stanford, CA: CSLI Publications.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Written and Spoken English.* London: Longman.

Birner, Betty & Gregory Ward. 1998. *Information Status and Noncanonical Word Order.* Amsterdam: John Benjamins.

Bresnan, Joan. 1977. Transformations and Categories in Syntax. In R. E. Butts & Jaakko Hintikka (eds.), *Basic Problems in Methodology and Linguistics. Part Three of the Proceedings of the Fifth International Congress of Logic, Methodology, and Philosophy of Science, London, Ontario, Canada*, 261–282. Dordrecht: Reidel.

Bresnan, Joan. 1978. A Realistic Transformational Grammar. In Morris Halle, Joan Bresnan & George A. Miller (eds.), *Linguistic Theory and Psychological Reality*, 1–59. Cambridge (MA): MIT Press.

Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2016. *Lexical-Functional Syntax.* Oxford/Malden (Mass.): Blackwell Publishers Ltd second edition edn.

Brown, William H. 1969. Method and Style in the Old English "Pastoral Care". *The Journal of English and Germanic Philology* 68(4). 666–684.

Büring, Daniel. 2007. Semantics, Intonation and Information Structure. In *The Oxford Handbook of Linguistic Interfaces*, 445–474. Oxford: Oxford University Press.

Chafe, Wallace. 1976. Givenness, Contrastiveness, Definiteness, Subjects, Topics and Point of View. In Charles N. Li (ed.), *Subject and Topic*, New York: Academic Press.

Choi, Hye-Won. 1999. *Optimizing Structure in Context: Scrambling and Information Structure*. Stanford, CA: CSLI Publications.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton de Gruyter.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge (MA): MIT Press.

Chomsky, Noam. 1981. *Lectures on Government and Binding* Studies in Generative Grammar 9. Dordrecht: Foris.

Colgrave, Betram & R. A. B. Mynors (eds.). 1969. *Bede's Ecclesiastical History of the English People*. Oxford: Clarendon Press.

Dalrymple, Mary. 2001. *Lexical Functional Grammar*, vol. 34 Syntax and Semantics. New York: Academic Press.

Dalrymple, Mary, Helge Dyvik & Tracy Holloway King. 2004. Copular Complements: Closed or Open. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG04 Conference*, Stanford, CA: CSLI Publications.

Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and Information Structure*. Cambridge (UK) / New York: Cambridge University Press.

Davies, William D. & Stanley Dubinsky. 2009. On the Existence (and Distribution) of Sentential Subjects. In D. B. Gerdts, J. C Moore & M. Polinksy (eds.), *Hypothesis A/Hypothesis B: Linguistic explorations in honor of David M. Perlmutter*, Cambridge (MA): MIT Press.

Delahunty, Gerald P. 1983. But Sentential Subjects Do Exist. *Linguistic Analysis* 12. 379–398.

Denison, David. 1993. *English Historical Syntax: Verbal Constructions*. London/New York: Longman.

Elmer, Willy. 1983. Semantic-Syntactic Patterning: The Lexical Valency of Seem in Middle English. *English Studies* 64(2). 160–168.

Emonds, Joseph E. 1976. *A Transformational Approach to English Syntax: Root, Structure-Preserving, and Local Transformations*. New York: Academic Press.

Engdahl, Elisabet. 2012. Optional Expletive Subjects in Swedish. *Nordic Journal of Linguistics* 35(2). 99–144.

Erdmann, Peter. 1988. On the Principle of 'Weight' in English. In Theo Venneman & Caroline Duncan-Rose (eds.), *On Language, Rhetorica Phonologica Syntactica: A Festschrift for Robert P. Stockwell from His Friends and Colleagues*, 325–339. London: Routledge.

Falk, Yehuda N. 2001. *Lexical-Functional Grammar: An Introduction to Parallel Constraint-Based Syntax.* Stanford, CA: CSLI Publication-sPublications.

Falk, Yehuda N. 2005. Open Argument Functions. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG05 Conference*, Stanford, CA: CSLI Publications.

Fanego, Teresa. 2004. Some Strategies for Coding Sentential Subjects in English. From Exaptation to Grammaticalization. *Studies in Language* 28(2). 321–361.

Fischer, Olga & Frederike Van Der Leek. 1983. The Demise of the Old English Impersonal Construction. *Journal of Linguistics* 19(2). 337–368.

Francis, Elaine J. 2010. Grammatical Weight and Relative Clause Extraposition in English. *Cognitive Linguistics* 21(1). 35–74.

van Gelderen, Elly. 2000. *A History of English Reflexive Pronouns: Person, Self and Interpretability.* Amsterdam: John Benjamins.

Gisborne, Nikolas & Jasper Holmes. 2007. A History of English Evidential Verbs of Appearance. *English Language and Linguistics* 11(1). 1–29.

Givón, Talmy. 2001. *Syntax. An Introduction.*, vol. 1. Amsterdam/Philadelphia: John Benjamins.

Gries, Stefan Th. 2009. *Statistics for Linguistics with R. A Practical Introduction.* Berlin: De Gruyter Mouton.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69(2). 274–307.

Haugland, Kari E. 2006. *Old English Impersonal Constructions and the Use and Non-Use of Nonreferential Pronouns.* Bergen: Universitetet i Bergen dissertation.

Hawkins, John A. 2004. *Efficiency and Complexity in Grammars.* Oxford/New York: Oxford University Press.

Horobin, Simon & Jeremy Smith. 2002. *An Introduction to Middle English*. Oxford/New York: Oxford University Press.

Huddleston, Rodney D. & Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Jespersen, Otto. 1909-1949. *A Modern English Grammar on Historical Principles*. Copenhagen: George Allen and Unwin Ltd/Ejnar Munksgaard.

Kaltenböck, Günther. 1999. Which *It* is It? Some Remarks on Anticipatory *It*. *VIEWS: Vienna English Working Papers* 8(2). 48–71.

Kaltenböck, Günther. 2005. It-Extraposition in English. A Functional View. *International Journal of Corpus Linguistics* 10(2). 119–159.

Kaplan, Ronald M. & Joan Bresnan. 1982. Lexical-Functional Grammar: A Formal System for Grammatical Representation. In Joan Bresnan (ed.), *The Mental Representation of Grammatical Relations*, 173–281. Cambridge (MA): MIT Press.

Karlsson, Fred. 2007. Constraints on Multiple Center-Embedding of Clauses. *Journal of Linguistics* 43(2). 365–392.

Katz, Jonah & Elisabeth Selkirk. 2011. Contrastive Focus vs. Discourse-New: Evidence from Phonetic Prominence in English. *Language* 87(4). 771–816.

Kibort, Anna. 2007. Extending the Applicability of Lexical Mapping Theory. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG07 Conference*, Stanford, CA: CSLI Publications.

Kibort, Anna. 2008. On the Syntax of Ditransitive Constructions. In Miriam Butt & Tracy Holloway King (eds.), *The Proceedings of the LFG '08 conference*, 312–332. Stanford, CA.

Kibort, Anna. 2013. Objects and Lexical Mapping Theory. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG13 Conference*, Stanford, CA: CSLI Publications.

Kibort, Anna. 2014. Mapping out a Construction Inventory with (Lexical) Mapping Theory. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG14 Conference*, Stanford, CA: CSLI Publications.

King, Tracy Holloway. 1997. Focus Domains and Information-Structure. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG97 Conference*, Stanford, CA: CSLI Publications.

Komen, Erwin. 2013. Lemma Lists for Old English and Middle English. http://erwinkomen.ruhosting.nl/eng/.

Koster, Jan. 1978. Why Subject Sentences Don't Exist. In S. J. Keyser (ed.), *Recent Transformational Studies in European Languages*, Cambridge (MA): MIT Press.

Krifka, Manfred & Renate Musan. 2012. Information Structure: Overview and Linguistic Issues. In Manfred Krifka & Renate Musan (eds.), *The Expression of Information Structure*, Berlin/Boston: De Gruyter Mouton.

Kroch, Anthony, Beatrice Santorini & Lauren Delfs. 2005. Penn-Helsinki Parsed Corpus of Early Modern English. CD-ROM, first edition, (http://www.ling.upenn.edu/hist-corpora/).

Kroch, Anthony & Ann Taylor. 2000. The Penn-Helsinki Parsed Corpus of Middle English. CD-ROM, second edition, (http://www.ling.upenn.edu/hist-corpora/).

Kroch, Anthony S., Beatrice Santorini & Ariel Diertani. 2010. Penn Parsed Corpus of Modern British English. CD-ROM, first edition, (http://www.ling.upenn.edu/hist-corpora/).

Lapidge, Michael. 1986. The Anglo-Latin Background. In Stanly B. Greenfield & David G. Calder (eds.), *A New Critical History of Old English Literature*, New York: New York University Press.

Lasnik, Howard. 2003. On the Extended Projection Principle. *Studies in Modern Grammar* 31. 1–23.

Light, Caitlin. 2012. *The Syntax and Pragmatics of Fronting in Germanic*. Philadelphia: University of Pennsylvania dissertation.

Lødrup, Helge. 2012. In Search of a Nominal COMP. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG12 Conference*, CSLI Publications.

Lohse, B., John A. Hawkins & Thomas Wasow. 2004. Domain Minimization in English Verb-Particle Constructions. *Language* 80. 238–261.

Los, Bettelou. 2005. *The Rise of the To-Infinitive*. Oxford/New York: Oxford University Press.

Méndez Naya, Belén. 1997. Subject Clauses in Old English: Do They Really Exist? *Miscelánea. A Journal of English and American Studies* 18. 213–230.

Miller, Philip H. 2001. Discourse Constraints on (Non)extraposition from Subject in English. *Linguistics* 39(4). 683–701.

Mitchell, Bruce. 1985a. *Old English syntax. Concord, the Parts of Speech, and the Sentence*, vol. I. Oxford: Clarendon Press.

Mitchell, Bruce. 1985b. *Old English Syntax: Subordination, Independent Elements, and Element Order*, vol. II. Oxford: Clarendon Press.

Nordlinger, Rachel & Joan Bresnan. 2011. Lexical-Functional Grammar: Interactions between Morphology and Syntax. In Robert Borsley & Kersti Börjars (eds.), *Non-Transformational Syntax: Formal and Explicit Models of Grammar*, Malden and Oxford: Blackwell Publishers Ltd.

Nordlinger, Rachel & Louisa Sadler. 2006. Verbless Clauses: Revealing the Structure within. In Jane Grimshaw, Joan Maling, Chris Manning, Jane Simpson & Annie Zaenan (eds.), *Architectures, Rules and Preferences: A Festschrift for Joan Bresnan*, Stanford, CA: CSLI Publications.

Pei, M. & F. Gaynor. 1954. *A Dictionary of Linguistics*. New York: Philosophical Library.

Postal, Paul M. 1974. *On Raising. One Rule of English Grammar and Its Theoretical Implications*. Cambridge (MA): The MIT Press.

Prince, Ellen. 1988. The ZPG Letter: Subjects, Definiteness, and Information-status. In William C. Mann & Sandra A. Thompson (eds.), *Discourse Description: Diverse linguistic analyses of a fund-raising text*, 295–325. John Benjamins.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London/New York: Longman.

Radford, Andrew. 2004. *Minimalist Syntax: Exploring the Structure of English*. Cambridge (UK): Cambridge University Press.

Ramhöj, Rickard. 2015. Clausal Subjects and Extraposition in the History of English. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG15 Conference*, Stanford, CA: CSLI Publications.

Randall, Beth. 2005-2007. *CorpusSearch 2*. Philadelphia: University of Pennsylvania.

Robertson, A. J. (ed.). 1956. *Anglo-Saxon Charters*. Cambridge (UK): Cambridge University Press.

Rosenbaum, Peter S. 1967. *The Grammar of English Predicate Complement Constructions*. Cambridge (MA): MIT Press.

Rusten, Kristian A. 2013. Empty Referential Subjects in Old English Prose: A Quantitative Analysis. *English Studies* 94(8). 970–992.

Seppänen, Aimo. 1986. The syntax of *seem* and *appear* revisited. *Studia Linguistica* 40(1). 22–39.

Seppänen, Aimo, Claes Göran Engström & Ruth Seppänen. 1990. On the So-Called Anticipatory *It*. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 43(6). 748–761.

Seppänen, Aimo & Jennifer Herriman. 2002. Extraposed Subjects vs. Postverbal Complements: On the So-Called Obligatory Extraposition. *Studia Neophilologica* 74(1). 30–59.

Shahar, Jed. 2008. *What Some* Its *are: Nonreferential* It*, Extraposition, and Copies*. Ann Arbor: City University of New York.

Stallings, Lynne M. & Maryellen C. MacDonald. 2011. It's not Just the "Heavy NP": Relative Phrase Length Modulates the Production of Heavy-NP Shift. *Journal of Psycholinguistic Research* 40. 177–187.

Stowell, Tim. 1981. *Origins of Phrase Structure*. Cambridge (MA): MIT dissertation.

Taylor, Ann, Arja Nurmi, Anthony Warner, Susan Pintzuk & Terttu Nevalainen. 2006. York-Helsinki Parsed Corpus of Early English Correspondence. Compiled by the CEEC Project Team. York: University of York and Helsinki: University of Helsinki. Distributed through the Oxford Text Archive.

Taylor, Ann, Anthony Warner, Susan Pintzuk & Frank Beths. 2003. The York-Toronto-Helsinki Parsed Corpus of Old English Prose. Oxford Text Archive, first edition, (http://www-users.york.ac.uk/ lang18/pcorpus.html).

Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Traugott, Elizabeth Closs. 1992. Syntax. In Richard Hogg (ed.), *The Cambridge History of the English Language*, vol. 1, 168–289. Cambridge University Press.

Treharne, Elaine. 2010. *Old and Middle English c.890-c.1450. An Anthology*. Malden and Oxford: Blackwell Publishers Ltd.

Vincent, Nigel. 2001. LFG as a Model of Syntactic Change. In Miriam Butt & Tracy Holloway King (eds.), *Time over Matter. Diachronic Perspectives on Morphosyntax*, Stanford, CA: CSLI Publications.

Visser, Fredericus Theodorus. 1963-1973. *An Historical Syntax of the English Language*. Leiden: E. J. Brill.

Wahlén, Nils. 1925. *The Old English Impersonalia*. Göteborg: Elanders Boktryckeri Aktiebolag.

Ward, Gregory & Betty Birner. 2004. Information Structure and Non-canonical Syntax. In Laurence Horn & Gregory Ward (eds.), *Handbook of Pragmatic Theory*, chap. 7. Blackwell Publishers Ltd.

Wasow, Thomas. 1997. Remarks on Grammatical Weight. *Language Variation and Change* 9. 81–105.

Zaenen, Annie & Elisabet Engdahl. 1994. Descriptive and Theoretical Syntax in the Lexicon. In B. T. S. Atkins & Antonio Zampolli (eds.), *Computational Approaches to the Lexicon*, 181–212. Oxford: Oxford University Press.

Zimmerman, Richard. 2015. Early English Clausal Arguments of Intransitive Verbs: Subjects or Associates of Empty Expletives? In *The York-Newcastle-Holland Symposium on the History of English Syntax 13 (SHES13)*, .

Appendix A

Definition and coding query files

A.1 Summary of definition file

The definition file (.def) is a file used for the CorpusSearch program to create shortcut commands for frequently used search strings. In the present section, the most important definitions are presented in the form they have in the .def file.

Frequently used syntactic concepts:

Sbj:	NP-NOM*|NP-SBJ*
Obj:	NP-ACC*|NP-OB1*
Obj2:	NP-DAT*|NP-OB2*

Examples from the lemma list, giving a list of forms and spelling variation for particular lexemes (based on Komen (2013)):

gelimpan: gelimpan|gelamp|gelomp|gelumpon|gelumpe|gelimpe+d|
gelimpa+d|gelimpe+t|gelimp+d|gelimpe|gelampt|gelumpan|
gelumpun|gelimpa+t|gelimp+t|gelymp+d|gelympe+d|
gelympelimpan|*limpende|gelumpen|*lamp|*lumpon|
*limp+d|*limp+d|*limpe+d|gelympan|*lamp|gelumpen|
*limpe+t|*limpe|*limpen|*lumpen|*lomp|*lymppede|
Gelimpan|Gelamp|Gelomp|Gelumpon|Gelumpe|
Gelimpe+d|Gelimpa+d|Gelimpe+t|Gelimp+d|Gelimpe|
Gelampt|Gelumpan|Gelumpun|Gelimpa+t|Gelimp+t|
Gelymp+d|Gelympe+d|Gelympe|Limpan|Gelumpen

thyncan: +tyncan|ge+tuht|+tincean|+tincan|+dincan|+dyncan|
+tuhte|+duhte|+duhton|ge+tuht|ge+duhte|+tync+d|+tinc+d|
+tinc+t|+dinc+d|+ting+d|+tinca+d|+dync+d|+dyncet|
+dyncea+d|+dynca+d|+tince+t|+tince+d|+dinca+d|+dince|
+tince|+dynce|+tynce|+dyncen|+dyncean|+tuhton|ge+duht|
ge+tuht|+dinc+t|+dynce+d|+tincea+d|+tynca+d|+tync+d|
+tynce+d|+tyncest|+tynce+t|ge+dynce|+Tyncan|Ge+tuht|
+Tincean|+Tincan|+Dincan|+Dyncan|+Tuhte|+Duhte|
+Duhton|Ge+tuht|Ge+duhte|+Tync+d|+Tinc+d|+Tinc+t|
+Dinc+d|+Ting+d|+Tinca+d|+Dync+d|+Dyncet|
+Dyncea+d|+Dynca+d|+Tince+t|+Tince+d|+Dinca+d|
+Dince|+Tince|+Dynce|+Tynce|+Dyncen|+Dyncean|
+Tuhton|Ge+duht|Ge+tuht|+Dinc+t|+Dynce+d|+Tincea+d|
+Tynca+d|+Tync+d|+Tynce+d|+Tyncest|+Tynce+t|
Ge+dynce

A.2 Coding queries

The coding queries are written down in a coding query file (.c), which can be used to label the structures found in the corpora according to the specifications of the query. The coding queries used for the present investigation are given below, in the same form they have in the .c file.

```
//Preamble commands
node: IP
coding_query:

//Periodization of texts
1: {
oe: (codocu1*|codocu2*|cobede*|coboeth*|cocura*|colaece*|colawaf*|colawafint*
|coorosiu*|coprefcura*|coalex*|coblick*|cochad*|cochronA*|codocu3*|codocu4*
|cogregdC*|cogregdH*|colacnu*|comart3*|comarvel*|coquadru*|coverhom*
|coalex*|coblick*|cochad*|cochronA*|codocu3*|codocu4*|cogregdC*|cogregdH*
|colacnu*|comart3*|comarvel*|coquadru*|coverhom*|coaelhom*|coaelive*
|coapollo*|cobenrul*|cobyrhtf*|cocathom1*|cocathom2*|codocu3*|coepigen*
|colaw1cn*|colaw2cn*|colaw5atr*|colaw6atr*|colawnorthu*|colsigef*|colwstan1*
|colwstan2*|cootest*|coprefcath1*|coprefcath2*|coprefgen*|copreflives*|cotempo*
|cowsgosp*|cogenesiC*|coherbar*|coeust*|cochronC*|cochronD*|coeuphr*
```

|cosevensl*|coverhomE*|coadrian*|cochronE*|codicts*|coinspolD*|colawger*
|colsigewZ*|colwsigeXa*|comargaC*|cowulf*|conicodA*|cochdrul*|covinsal*
|comart2*|colawwllad*|coleofri*|cosolsat1*|coalcuin*|coneot*|corood*|cochristoph*
|coprefsolilo*|cosolilo*|coeluc1*|comary*|conicodD*|conicodE*|coinspolX*|
cojames*|colsigeB*|colwgeat*|comargaT*|comary*|colsigewB* inID)
me: (*CMVICES1*|*CMVICES1*|*CMTRINIT*|*CMSAWLES*|*CMPETERB*
|*CMORM*|*CMMARGA*|*CMLAMBX1*|*CMLAMB1*|*CMKENTHO*
|*CMKATHE*|*CMJULIA*|*CMHALI*|*CMANCRIW*|*CMROLLTR*
|*CMROLLEP*|*CMKENTSE*|*CMEARLPS*|*CMAYENBI*|*CMAELR3*
|*CMWYCSER*|*CMVICES4*|*CMROYAL*|*CMPURVEY*|*CMPOLYCH*
|*CMOTEST*|*CMNTEST*|*CMMIRK*|*CMMANDEV*|*CMJULNOR*
|*CMHORSES*|*CMHILTON*|*CMGAYTRY*|*CMEQUATO*|*CMEDVERN*
|*CMEDTHOR*|*CMCTPARS*|*CMCTMELI*|*CMCLOUD*|*CMBRUT3*
|*CMBOETH*|*CMBENRUL*|*CMASTRO*|*CMTHORN*|*CMSIEGE*
|*CMREYNES*|*CMREYNAR*|*CMMALORY*|*CMKEMPE*|*CMINNOCE*
|*CMGREGOR*|*CMFITZJA*|*CMEDMUND*|*CMCAPSER*|*CMCAPCHR*
|*CMAELR4*|AUTHNEW*|AUTHOLD*|BOETHEL*|ERV-*|NEWCOME-*
|PURVER-*|TYNDNEW*|TYNDOLD*|CMAYEN*|CMBRUT3*|CMLAMB*
|*m*1,*|*m2*,*|*m3,*|*m*4,* inID)
emod: (*E1*|*E2*|*E3* inID)
mod: (*1700*|*1707*|*1710*|*1711*|*1712*|*1716*|*1718*|*1719*|*1726*
|*1736*|*1740*|*1742*|*1743*|*1744*|*1745*|*1746*|*1747*|*1749*|*1753*
|*1762*|*1763*|*1764*|*1769*|*1773*|*1774*|*1775*|*1776*|*1777*|*1780*
|*1785*|*1793*|*1796*|*1797*|*1799*|*1800*|*1805*|*1806*|*1807*|*1808*
|*1813*|*1814*|*1815*|*1817*|*1826*|*1830*|*1835*|*1836*|*1837*|*1859*
|*1861*|*1863*|*1865*|*1866*|*1873*|*1876*|*1878*|*1881*|*1882*|*1885*
|*1886*|*1890*|*1895*|*1897*|*1900*|*1901*|*1905*|*1908*|*1913*
|*-17[01234].*|*-17[56789].*|*-18[01234].*|*-18[56789].*|*-19[01].* inID)
z: ELSE }

//Number of IPs in the corpora:
2: { clause: (IP* exists)
z: ELSE }

//Constructions:
3: {nonextra: ((CP*SBJ* exists) OR (IP-INF-SBJ* exists)) AND (!CP-FRL-SBJ*
exists)
extraexp: (NP-NOM*|NP-SBJ* iDoms PRO*) AND ((NP-NOM*|NP-SBJ*
sameIndex IP-INF*|CP-THT*|CP-QUE*) OR (NP-NOM-x hasSister
IP-INF-x|CP-THT-x|CP-QUE-x)) AND (NP-NOM*|NP-SBJ* hasSister
!NP-MSR*|NP-2*|NP-1*|NP-NOM)
extranonexp: (NP-NOM*|NP-SBJ* iDoms exp) AND ((NP-NOM*|NP-SBJ*
sameIndex IP-INF*|CP-THT*|CP-QUE*) OR (NP-NOM-x hasSister
IP-INF-x|CP-THT-x|CP-QUE-x)) AND (NP-NOM*|NP-SBJ* hasSister
!NP-MSR*|NP-2*|NP-1*|NP-NOM)
z: ELSE }

//Clause types:
4: {that: (finite_verb hasSister CP-THT*)
wh: (finite_verb hasSister CP-QUE*)
inf: (finite_verb hasSister IP-INF*)
z: ELSE }

// Presence of oblique experiencer:
5: {obl: (V* hasSister NP-OB2*|NP-DAT*|NP) OR ((V* hasSister PP) AND (PP iDoms P) AND (P iDoms to|To))
nonobl: (V* hasSister !NP-OB2*|NP-DAT*|NP)
z: ELSE }

//Lexical verbs in Old English:
6: {thyncan: (V* iDoms thyncan)
gelimpan: (V* iDoms gelimpan)
z: ELSE }

//Latin original or not for the Old English texts
7: {latin: (*coalex*|*coapollo*|*cobede*|*cobenrul*|*cobyrhtf*|*cochad*
|*cochdrul*|*cocura*|*cocuraC*|*codicts*|*cogregdC*|*cogregdH*
|*coherbar*|*colacnu*|*colaece*|*colsigef*|*colwgeat*|*colwstan1*
|*colwstan2*|*comargaC*|*comarvel*|*conicodA*|*conicodC*|*conicodD*
|*conicodE*|*coorosiu*|*cootest*|*coquadru*|*corood*|*cosolilo*
|*cotempo*|*covinsal*|*cowsgosp* inID)
notrans: (*coadrian*|*coaelhom*|*coaelive*|*coaugust*|*coblick*
|*cocanedgD*|*cocanedgX*|*cocathom1*|*cocathom2*|*cochronA*
|*coepigen*|*coeuphr*|*coeust*|*coinspolD*|*coinspolX*|*colaw1cn*
|*colaw2cn*|*colaw5atr*|*colaw6atr*|*colawaf*|*colawafint*|*colawger*
|*colawine*|*colawnorthu*|*colawwllad*|*coleofri*|*colsigewZ*
|*colwsigeXa*|*comart1**comart2*|*comart3*|*comary*|*coprefcath1*
|*coprefcath2*|*coprefcura*|*coprefgen*|*copreflives*|*coprefsolilo*
|*cosevensl*|*cosolsat1*|*cosolsat2*|*cowulf* inID)
unknown: (*coalcuin*|*coboeth*|*cochristoph*|*cochronC*|*cochronD*
|*cochronE*|*coducu1*|*coducu2*|*coducu2*|*coducu3*|*coducu3*
|*coducu4*|*coeluc1*|*coeluc2*|*coexodusP*|*cogenesiC*|*cojames*
|*colsigewB*|*colwsigeT*|*comargaT*|*coneot*|*coverhom*
|*coverhomE*|*coverhomL*|*covinceB* inID)
z: ELSE }

//Number of words of sentence:
8: {\2: (IP-MAT*|IP-SUB* DomsWords 2)
\3: (IP-MAT*|IP-SUB* DomsWords 3)
\4: (IP-MAT*|IP-SUB* DomsWords 4)
\5: (IP-MAT*|IP-SUB* DomsWords 5)
\6: (IP-MAT*|IP-SUB* DomsWords 6)
\7: (IP-MAT*|IP-SUB* DomsWords 7)
\8: (IP-MAT*|IP-SUB* DomsWords 8)
\9: (IP-MAT*|IP-SUB* DomsWords 9)
\10: (IP-MAT*|IP-SUB* DomsWords 10)

\11: (IP-MAT*|IP-SUB* DomsWords 11)
\12: (IP-MAT*|IP-SUB* DomsWords 12)
\13: (IP-MAT*|IP-SUB* DomsWords 13)
\14: (IP-MAT*|IP-SUB* DomsWords 14)
\15: (IP-MAT*|IP-SUB* DomsWords 15)
\16: (IP-MAT*|IP-SUB* DomsWords 16)
\17: (IP-MAT*|IP-SUB* DomsWords 17)
\18: (IP-MAT*|IP-SUB* DomsWords 18)
\19: (IP-MAT*|IP-SUB* DomsWords 19)
\20: (IP-MAT*|IP-SUB* DomsWords 20)
\21: (IP-MAT*|IP-SUB* DomsWords 21)
\22: (IP-MAT*|IP-SUB* DomsWords 22)
\23: (IP-MAT*|IP-SUB* DomsWords 23)
\24: (IP-MAT*|IP-SUB* DomsWords 24)
\25: (IP-MAT*|IP-SUB* DomsWords 25)
\26: (IP-MAT*|IP-SUB* DomsWords 26)
\27: (IP-MAT*|IP-SUB* DomsWords 27)
\28: (IP-MAT*|IP-SUB* DomsWords 28)
\29: (IP-MAT*|IP-SUB* DomsWords 29)
\30: (IP-MAT*|IP-SUB* DomsWords 30)
\31: (IP-MAT*|IP-SUB* DomsWords 31)
\32: (IP-MAT*|IP-SUB* DomsWords 32)
\33: (IP-MAT*|IP-SUB* DomsWords 33)
\34: (IP-MAT*|IP-SUB* DomsWords 34)
\35: (IP-MAT*|IP-SUB* DomsWords 35)
\36: (IP-MAT*|IP-SUB* DomsWords 36)
\37: (IP-MAT*|IP-SUB* DomsWords 37)
\38: (IP-MAT*|IP-SUB* DomsWords 38)
\39: (IP-MAT*|IP-SUB* DomsWords 39)
\40: (IP-MAT*|IP-SUB* DomsWords 40)
\41: (IP-MAT*|IP-SUB* DomsWords 41)
\42: (IP-MAT*|IP-SUB* DomsWords 42)
\43: (IP-MAT*|IP-SUB* DomsWords 43)
\44: (IP-MAT*|IP-SUB* DomsWords 44)
\45: (IP-MAT*|IP-SUB* DomsWords 45)
\46: (IP-MAT*|IP-SUB* DomsWords 46)
\47: (IP-MAT*|IP-SUB* DomsWords 47)
\48: (IP-MAT*|IP-SUB* DomsWords 48)
\49: (IP-MAT*|IP-SUB* DomsWords 49)
\50: (IP-MAT*|IP-SUB* DomsWords 50)
\51: (IP-MAT*|IP-SUB* DomsWords 51)
\52: (IP-MAT*|IP-SUB* DomsWords 52)
\53: (IP-MAT*|IP-SUB* DomsWords 53)
\54: (IP-MAT*|IP-SUB* DomsWords 54)
\55: (IP-MAT*|IP-SUB* DomsWords 55)
\56: (IP-MAT*|IP-SUB* DomsWords 56)
\57: (IP-MAT*|IP-SUB* DomsWords 57)
\58: (IP-MAT*|IP-SUB* DomsWords 58)
\59: (IP-MAT*|IP-SUB* DomsWords 59)
\60: (IP-MAT*|IP-SUB* DomsWords 60)

z: ELSE }

//Number of words of clausal complement
9: {\2: (IP-INF*|CP-THT*|CP-QUE* DomsWords 2)
\3: (IP-INF*|CP-THT*|CP-QUE* DomsWords 3)
\4: (IP-INF*|CP-THT*|CP-QUE* DomsWords 4)
\5: (IP-INF*|CP-THT*|CP-QUE* DomsWords 5)
\6: (IP-INF*|CP-THT*|CP-QUE* DomsWords 6)
\7: (IP-INF*|CP-THT*|CP-QUE* DomsWords 7)
\8: (IP-INF*|CP-THT*|CP-QUE* DomsWords 8)
\9: (IP-INF*|CP-THT*|CP-QUE* DomsWords 9)
\10: (IP-INF*|CP-THT*|CP-QUE* DomsWords 10)
\11: (IP-INF*|CP-THT*|CP-QUE* DomsWords 11)
\12: (IP-INF*|CP-THT*|CP-QUE* DomsWords 12)
\13: (IP-INF*|CP-THT*|CP-QUE* DomsWords 13)
\14: (IP-INF*|CP-THT*|CP-QUE* DomsWords 14)
\15: (IP-INF*|CP-THT*|CP-QUE* DomsWords 15)
\16: (IP-INF*|CP-THT*|CP-QUE* DomsWords 16)
\17: (IP-INF*|CP-THT*|CP-QUE* DomsWords 17)
\18: (IP-INF*|CP-THT*|CP-QUE* DomsWords 18)
\19: (IP-INF*|CP-THT*|CP-QUE* DomsWords 19)
\20: (IP-INF*|CP-THT*|CP-QUE* DomsWords 20)
\21: (IP-INF*|CP-THT*|CP-QUE* DomsWords 21)
\22: (IP-INF*|CP-THT*|CP-QUE* DomsWords 22)
\23: (IP-INF*|CP-THT*|CP-QUE* DomsWords 23)
\24: (IP-INF*|CP-THT*|CP-QUE* DomsWords 24)
\25: (IP-INF*|CP-THT*|CP-QUE* DomsWords 25)
\26: (IP-INF*|CP-THT*|CP-QUE* DomsWords 26)
\27: (IP-INF*|CP-THT*|CP-QUE* DomsWords 27)
\28: (IP-INF*|CP-THT*|CP-QUE* DomsWords 28)
\29: (IP-INF*|CP-THT*|CP-QUE* DomsWords 29)
\30: (IP-INF*|CP-THT*|CP-QUE* DomsWords 30)
\31: (IP-INF*|CP-THT*|CP-QUE* DomsWords 31)
\32: (IP-INF*|CP-THT*|CP-QUE* DomsWords 32)
\33: (IP-INF*|CP-THT*|CP-QUE* DomsWords 33)
\34: (IP-INF*|CP-THT*|CP-QUE* DomsWords 34)
\35: (IP-INF*|CP-THT*|CP-QUE* DomsWords 35)
\36: (IP-INF*|CP-THT*|CP-QUE* DomsWords 36)
\37: (IP-INF*|CP-THT*|CP-QUE* DomsWords 37)
\38: (IP-INF*|CP-THT*|CP-QUE* DomsWords 38)
\39: (IP-INF*|CP-THT*|CP-QUE* DomsWords 39)
\40: (IP-INF*|CP-THT*|CP-QUE* DomsWords 40)
\41: (IP-INF*|CP-THT*|CP-QUE* DomsWords 41)
\42: (IP-INF*|CP-THT*|CP-QUE* DomsWords 42)
\43: (IP-INF*|CP-THT*|CP-QUE* DomsWords 43)
\44: (IP-INF*|CP-THT*|CP-QUE* DomsWords 44)
\45: (IP-INF*|CP-THT*|CP-QUE* DomsWords 45)
\46: (IP-INF*|CP-THT*|CP-QUE* DomsWords 46)
\47: (IP-INF*|CP-THT*|CP-QUE* DomsWords 47)
\48: (IP-INF*|CP-THT*|CP-QUE* DomsWords 48)

\49: (IP-INF*|CP-THT*|CP-QUE* DomsWords 49)
\50: (IP-INF*|CP-THT*|CP-QUE* DomsWords 50)
\51: (IP-INF*|CP-THT*|CP-QUE* DomsWords 51)
\52: (IP-INF*|CP-THT*|CP-QUE* DomsWords 52)
\53: (IP-INF*|CP-THT*|CP-QUE* DomsWords 53)
\54: (IP-INF*|CP-THT*|CP-QUE* DomsWords 54)
\55: (IP-INF*|CP-THT*|CP-QUE* DomsWords 55)
\56: (IP-INF*|CP-THT*|CP-QUE* DomsWords 56)
\57: (IP-INF*|CP-THT*|CP-QUE* DomsWords 57)
\58: (IP-INF*|CP-THT*|CP-QUE* DomsWords 58)
\59: (IP-INF*|CP-THT*|CP-QUE* DomsWords 59)
\60: (IP-INF*|CP-THT*|CP-QUE* DomsWords 60)
z: ELSE }

//Givenness status of subclause
10: {pro: (IP-INF*|CP-THT*|CP-QUE* Doms NP*) AND (NP* iDoms PRO*)
det: (IP-INF*|CP-THT*|CP-QUE* Doms NP*) AND (NP* iDoms D*) AND (D* iDoms !a*)
such: (IP-INF*|CP-THT*|CP-QUE* Doms NP*) AND (NP* iDoms SUCH)
z: ELSE }