

Variability in verb complementation in Late Modern English: Finite vs. non--finite patterns

HUBERT CUYCKENS, FRAUKE D'HOEDT & BENEDIKT SZMRECSANYI

10.1 Introduction¹

In Present-Day English (PDE), the following alternation pattern with the verb

remember is commonly attested:

- (1) a. *He remembered to thank her for everything.*
- b. *I remember reading about it in the newspaper.* (Declerck 1991: 511)

As can be seen, the verb *remember* may take a *to*-infinitive clause (1a) or a gerundial *-ing*-clause (1b) as clausal verb complement. Language users cannot choose freely between one or the other pattern; rather, the variation can be characterized as functional differentiation, whereby the *to*-infinitive encodes a situation which is not

¹ The research for this chapter was carried out during the first author's research stay at the English Department of the University of Santiago de Compostela. He would like to thank the Research Foundation Flanders (FWO) and the University of Leuven for granting, and financially supporting, his sabbatical leave. He also acknowledges the financial support of the Spanish Ministry for Science and Innovation and the European Regional Development Fund (grant HUM2007-60706/FILO, coordinated by Teresa Fanego). All authors are thankful for useful suggestions from the audience at the workshop on *Late Modern English Syntax*, organized by Marianne Hundt, University of Zurich, 19–20 August 2012, and for careful feedback from the volume editor and an anonymous reviewer. The usual disclaimers apply.

yet actualized at the time of remembering, whereas the gerundial *-ing*-clause encodes the situation as actualizing before or at that time. However, in the alternation pattern in (2) with *regret* (*that*-clause vs. gerundial *-ing*-clause), preference for one or the other verbal complement is far less categorical; actually, it would appear that both variants are freely interchangeable.

- (2) a. *I don't regret helping her start out.*
b. *I don't regret that I helped her start out.*

It is this type of *non-categorical* or *probabilistic* variation, which is less well understood, that is the focus of this chapter. Our aim is to offer a corpus-based analysis of complement-clause variation (in particular, finite *that*-clauses alternating with non-finite gerundial *-ing*-clauses and *to*-infinitive clauses) with the complement-taking predicates *remember*, *regret*, and *deny*. Focusing on the LModE (Late Modern English) period, we detail the changing patterns of complement choice over time. We also identify the various (changing) factors determining this complement clause variation, thus probing the multivariate and probabilistic nature of complement choice. Another aim is to examine to what extent changing/varying distributions of finite vs. non-finite complement clauses can inform more general hypotheses about complement choice.

The chapter is organized as follows: section 10.2 presents a short survey of clausal verb complementation, which serves as a backdrop to our study of complement clause variation. Section 10.3 details the goals of this chapter, while section 10.4 presents data and methodology. Section 10.5 is devoted to the multivariate and probabilistic

analysis of finite vs. non-finite complement choice. In section 10.6, we contrast the results of our analysis with earlier claims in the literature on (changing) complement-clause preferences. Section 10.7 summarizes and details the contribution our study makes to the body of previous research.

10.2 A short survey of clausal verb complementation

Clausal verb complementation (i.e., structures of the type [complement-taking predicate + *to*-infinitive clause/gerundial *-ing*-clause/*that*-clause/*for...to*-infinitive clause]) has been an important research topic within generative as well as cognitive-functional linguistic frameworks. Indeed, complementation phenomena have been a concern of generative linguists (Rosenbaum 1967, Bresnan 1970) since Chomsky's seminal *Aspects of a Theory of Syntax* (1965), and important work in this domain has continued ever since (Bresnan 1979, Warner 1982, Chomsky 1986, Rizzi 1990, Radford 1997, Felser 1999). From the 1980s onwards, the scope of complementation research was expanded by functional-typological linguists (e.g., Givón 1980, Noonan 1985, Dixon 1991). While the generative literature on verb complementation has largely focused on synchronic, syntactic issues, such as the constituent structure of different complement types, synchronic research within the cognitive-functional tradition to date has been concerned with the question of how complement clauses are distributed over the various complement-taking predicates (or main/matrix verbs). Much of this cognitive-functional research is semantic in orientation, and has been informed by Noonan's observation that across languages "complementation is basically a matter of matching a particular complement type to a particular complement-taking predicate" (1985: 90) (see Wierzbicka 1988, Duffley 1992, 1999, Langacker 1991, Achard 1998, Smith 2002). However, as De Smet (2013: 20–33)

points out, a satisfactory synchronic account of complementation needs to envisage additional, non-semantic determining principles, such as the role of information structure (Noël 2003), the *horror aequi* principle (Rudanko 2000, Vosberg 2003), the cognitive complexity principle (Rohdenburg 1995), social and regional stratification (Mair 2002, 2003), and register (Mindt 2000).

Significantly, though, this synchronic work has tended to neglect the fact that the synchronic matches between complement-taking predicate (CTP) and complement clause (CC) can be subject to change over time. It is only in the last fifteen or twenty years that diachronic studies have appeared which present broad accounts of change and variation in complementation patterns in different periods of the history of English (see Fischer 1995, Fanego 1996, 1998, Rudanko 1998, 2006, 2010, 2012, Miller 2001, Los 2005, Rohdenburg 2006a, De Smet 2008, 2013), attesting to a distributional reorganization of CCs over time.² It has thus been shown that new complement types have made their way into the English language (e.g., the gerundial *-ing*-CC and the *for...to*-infinitive construction are relatively recent), or that long existing types have spread to new CTP-contexts (e.g., the *to*-infinitive) or, conversely, have become increasingly restricted (e.g., the bare infinitive). These changes have led to competition between CC-types; in particular, in some CTP-contexts, one CC-type may have been replaced by another (e.g., with verbs of volition, the *that*-clause has largely been supplanted by the *to*-infinitive; see Croft 2000, Los 2005); in other contexts, a situation of variation may have come about. Sometimes, as in (1), this CC-variation results in functional differentiation, whereby the domain of operation of one

² Most linguists, also of the generative persuasion, are now agreed that the diachronic perspective is an important facet of a satisfactory account of complementation.

CC-type can be differentiated in a clear-cut, i.e., categorical, fashion from that of the other. Often, though, this variation is non-categorical, and unstable, in that CC-types co-exist (e.g., *regret* patterns with a *that*-clause, a gerundial *-ing*-clause, and a *to*-infinitive; verbs of ‘like’ and ‘love’ alternate between the *to*-infinitive and the gerundial *-ing* form) and show varying/shifting patterns of preference across speakers and over time. While replacement of CC-types as well as CC-variation characterized by functional differentiation have been well documented, the type of competition between CC-types resulting in unstable variation (or changing patterns of preference over time) – though attested in the literature – is still largely underexplored, and it is this non-categorical variation that this chapter is concerned with.

10.3 Goals

In this chapter we wish to examine non-categorical or probabilistic variation with the verbs *remember*, *regret*, and *deny*, as exemplified in (2) above and in (3):

- (3) a. *I do not remember my saying I knew the thief.* (Old Bailey Corpus (OBC), t17640222–37)
- b. *I don't remember I said your Evidence would go before Askin's.* (OBC, t17440912–48)

What characterizes this type of variation is indeterminacy of choice; this indeterminacy does not occur at the abstract level of grammar – indeed, each of the CC-types with *remember* in (3) can be independently motivated (see, for instance, the semantic characterization of CCs in Langacker 2008: 429–445) – but at the usage level, i.e., at the level of actual, online choices speakers make in discourse (see De

Smet 2013: 27–29). Our concern is with these online choices: (i) we explore the patterns of CC-preference (in terms of frequency distributions) that emerge from actual usage, as observed in corpora, and (ii) by fitting a logistic regression model, we aim to isolate the factors significantly predicting, in terms of odds or probability ratios, this CC-variation.

Finally, the frequency distributions observed and, importantly, the factors that turn out to be instrumental in predicting CC-choice feed into an assessment of (some of the) earlier claims made in the literature on (changing patterns in) CC-choice (see section 10.2; De Smet 2013: 20–33). Thus, we are able to explore to what extent CC-choice is affected (i) by Noonan’s semantic characterization of CCs in terms of ‘dependent time reference’ vs. ‘independent time reference’ (1985: 92), and (ii) by various types of structural complexity consonant with Rohdenburg’s (1995, 1996) notion of ‘cognitive complexity’. Further, we examine whether any changes in CC-choice observed over time tie in with Denison’s claim that “a long-term trend in English has been the growth of nonfinite complement clauses at the expense of finite clauses” (1998: 256), and whether any such increase of non-finite complements can inform our views on grammaticalization processes.

10.4 Data and methodology

10.4.1 Data selection

In our selection of matrix verbs, we chose verbs that allow variation between finite and non-finite complement clauses, and these are relatively few in number. We selected the verbs *remember*, *regret*, and *deny* from Quirk et al.’s (1985: 1182–1184)

list of factual verbs which allow the finite/non-finite CC alternation.³ Finite complement clauses are introduced either by the complementizer *that* or by ‘zero’, as in (4a) and (4b), respectively:

- (4) a. *I remember perfectly well that it was at the prisoner’s suggestion.*
(OBC, t-18720226–267)
- b. *I remember Boswell and Ausser were both at my house.* (OBC, t17670715–45)

Non-finite complement clauses include (i) the subjectless (controlled) *-ing*-CC (as in (5a)), (ii) the *-ing*-CC with expressed subject (5b), (iii) the subjectless *to*-infinitive (5c); (iv) the *to*-infinitive-CC with expressed subject (5d). In addition, care was taken that matrix verbs occurred sufficiently frequently with the CC-variants to allow statistically significant results.

- (5) a. *Do you remember at any time going to the prisoner’s house?* (OBC, t17730217–52)
- b. *Do you remember a green cart coming up?* (OBC, t17970920–62)
- c. *I do not remember ever to have heard a word from you before.* (1810–1813, Corpus of Late Modern English Texts (extended version) (CLMETEV), George Byron, *Letters*)

³ Quirk et al.’s (1985) list also includes *suasive* and *emotive* verbs, which will not be considered here. Even with the *suasive* and *emotive* verbs, the number of verbs allowing both finite and non-finite CCs remains limited.

- d. *I do remember this circumstance to have happened but to one man.*
(OBC, unit id = t17950218–46)

Data were extracted from two corpora covering the entire LModE period: the *Old Bailey Corpus* (OBC, Huber et al. 2012), which contains court transcripts and is therefore ‘speech-based’ (see Culpeper and Kytö 2010: 16–17)⁴ and the *Corpus of Late Modern English Texts (extended version)* (CLMETEV), which mainly contains formal prose.⁵

We extracted all attestations of the matrix verbs (including those with NP complements). After manual pruning of all spurious hits, we were left with the following number of [CTP – CC] patterns:

- with *remember*: 4371 observations
- with *regret*: 280 observations
- with *deny*: 1138 observations

⁴ Culpeper and Kytö (2010: 17) classify trial proceedings as ‘speech-based’:

“‘Speech-based’ genres, such as Trial proceedings, are those that are based on an actual ‘real-life’ speech event. There is no claim here that such genres involve the accurate recording of a speech event. In the absence of audio or video recording equipment or even full systems of shorthand, most speech-based texts are reconstructions assisted by notes.”

⁵ A new version of the corpus, CLMET 3.0, was not yet available when the research for this chapter was carried out (see

<https://perswww.kuleuven.be/~u0044428/clmet.htm>).

Observations were then coded, as described below.

10.4.2 Coding of the data

Each corpus attestation, consisting of [subject + *remember/regret/deny* + CC], was entered into an Excel database under the descriptor OBSERVATION.⁶ Each token was then coded for a number of factors determining CC-choice and describing characteristics of the complement-taking predicate/matrix verb, the complement clause, or the combined [CTP – CC] structure. Importantly, our selection of potentially significant factors of CC-variation was informed by the relevant literature (see section 10.2; De Smet 2013: 20–33) and comprises semantic, structural, genre-related, and periodization-related factors; it also includes a number of factors not typically discussed in the literature.^{7,8}

⁶ Each observation received the following additional descriptors: (i) N: number of the attestation in the Excel table; (ii) CONCORDANCE: the entire concordance comprising the corpus attestation (in order to provide the necessary context); (iii) LEMMA: *remember, regret, or deny*; (iv) YEAR: year of the attested example; (v) FILE: corpus in which the example was attested.

⁷ Information-based factors (as discussed in Noel 2003) were not included, as the present study focuses on [CTP – CC] patterns only. Nor was *horror aequi* (Rudanko 2000, Vosberg 2003) or factors relating to geographical or social stratification (Mair 2002, 2003).

⁸ The information in SMALL CAPS indicates the factor, as it was labeled in the Excel database.

The following semantic factors were coded for: (i) the meaning of the matrix clause, (ii) the meaning of the complement clause, (iii) [CTP – CC] TIME REFERENCE.⁹

- (i) *The meaning of the matrix verb* (VERB MEANING). The values making up this factor are distinguished at the level of the lemma's submeanings; not all submeanings of a verb will show the same CC-pattern.¹⁰

For *remember*, three meanings/values were distinguished:

- 'remember1' = 'recall', 'bring to mind'. This meaning is typically exemplified in (6a), where the subject recalls a past event/action/situation. In addition, one can also recall/bring to mind something that is a general truth or a situation still holding at the time

⁹ Our set of semantic factors potentially determining CC-choice does not include such dichotomies as 'internal' vs. 'holistic' (meant to capture the variation between gerundial *-ing*-clauses vs. *to*-infinitives; see Smith 2002; Langacker 2008: 439) or 'occurrence/process' vs. 'proposition/grounded process' (meant to capture the variation between non-finite and finite CCs; see Langacker 2008: 441–442). These dichotomies are meant to account for the distribution of CCs in a categorical fashion. Moreover, they are very general, and therefore hard to operationalize; it would seem, then, that they only play a minor role, if any, in speakers' on-line choice of CCs.

¹⁰ Careful inspection of the database of [CTP – CC] tokens shows that not all meanings of the matrix verbs allow finite vs. non-finite variation. As such, in our analysis, we will not consider the tokens with semantic values 'remember2' and 'remember3', which exclusively take a *to*-infinitive and *that*-clause, respectively. We will return later to these sets (section 10.6), and try to account for the fact that *remember* in these meanings exclusively takes one type of CC only.

of the matrix verb, as in (6b); or one can recall that one still has to do something, as in (6c):

- ‘remember2’ = ‘not to forget to do something’, as in (7):
- ‘remember3’ = ‘keep in mind’, as in (8):

- (6) a. *I remember a detective coming and speaking to me.* (OBC, t-18811212–125)
- b. *The Prisoner said in his Defence, that coming in a Chaise to the Vine Tavern, and remembering his Pistols were Charged, he went to discharge them ...* (OBC, t17211206–42)
- c. *(All of a sudden), he remembered that he (still) had to do the dishes.*
(non-attested example)¹¹
- (7) *Remember to lay your Spitchcot Eels near the edge of the Dish.* (1732, CLMETEV, Richard Bradley, *The Country Housewife and Lady Director*)
- (8) a. *Remember, you are upon your oath.* (OBC, t17980704–44)
- b. *remember he is come of gentle blood, ...* (1824, CLMETEV, Catherine Gore, *Theresa Marchont*)

For *regret*, two meanings were distinguished:

- ‘regret1’ = ‘to feel sorry about something one has done and that one should have done differently or about a state of affairs one is involved in or responsible for and that one wishes was different’, as in (9):

¹¹ This meaning was not attested in our corpus examples, but included here for completeness’ sake.

- 'regret2' is a more 'polite' use of *regret* where the speaker says that he/she is sorry or sad about a situation, usually one that he/she is not directly responsible for, as in (10):

- (9) a. *She regretted having made the appointment.* (1891, CLMETEV, George Gissing, *New Grub Street*)
b. *He seemed to regret that he had not finished the boy.* (OBC, t18980913–621)
- (10) a. *I much regret your leaving my office in such a hurry yesterday.* (OBC, t18960323–332)
b. *We regret the envelopes are not the same as per specimen sent you.* (OBC, t19100426–40)

Deny has been interpreted as follows: 'to say that you did not do something that someone has accused you of doing, or to say that something is not true'.

- (11) a. *He denied being charged with robbery.* (OBC, t-18680504–460)
b. *He denied that any bad money was found upon him.* (OBC, t-18730407–314)
- (ii) *Meaning of the complement clause* (MEANING CC). The complement clause may denote a state, as in (12), or an event/action as in (13).

- (12) *She had very much regretted being from home.* (1811, CLMETEV, Jane Austen, *Sense and Sensibility*)
- (13) *In short, to sum you up, you have all the makings in you of an ideal wife combined with faults sufficient to make a Socrates regret he'd ever married you.* (1909, CLMETEV, Jerome K. Jerome, *They and I*)

(iii) [CTP – CC] TIME-REFERENCE. This factor designates the relation between the time reference of the complement and the meaning of the CTP, and has the following values:

- ITR (independent time reference) indicates that “the time reference of the complement ... is in no way logically bound by the time reference of the CTP” (Noonan 1985: 92); as such, *deny* in (14) combines with CTPs with anterior (14a) as well as simultaneous time reference (14b):

- (14) a. *Prisoner denied that he knowingly bought stoled Goods.* (OBC, t17210301–56)
- b. *He absolutely denied that he knew any thing of the matter.* (OBC, t17350911–14)

- DTR (dependent time reference) indicates that the time reference of the complement “is a necessary consequence of the meaning of the CTP” (Noonan 1985: 92), as in (15):

- (15) *Remember to keep your Fire gentle and clear.* (1732, CLMETEV, Richard Bradley, *The Country Housewife and Lady’s Director*)

The following set of factors pertain to structural complexity:

- (iv) TYPE SUBJECT CC. This factor has the following two values:
 - ‘complex NP’, comprising [noun + postmodifier] and [noun + noun]
 - ‘other’, comprising the following types of subject: pronoun, noun, *there*, (pro)noun in the genitive, proper name, and no subject (when no subject is expressed in control environments)

- (v) COMPLEXITY CC. The CC may show the following types of predicate structure:
 - verbs without argument/modifier: ‘V’
 - verbs with one argument/modifier: ‘Vobjadv’
 - verbs with an argument + modifier, or two arguments, or two modifiers: ‘Vobj+adv’

- (vi) INTERVENING MATERIAL IN WORDS. Indicates the number of words between the CTP and the subject of the complement (when the complement is finite) or between the subject and the non-finite verb form.

The following genre-related factor was distinguished:

- (vii) MEDIUM. The attestation belongs to the ‘speech-based’ or the ‘written’ genre. Attestations from the CLMTEV, which contains formal prose, were classified as ‘written’, and attestations from the OBC as ‘speech-based’.

Periodization is captured by the following factor:

(viii) PERIOD. This factor locates the attestation within two time bands of the LModE period:

- early (between 1710 and 1780)
- late (between 1781 and 1920)

We first experimented with three 70-year time bands (1710–1780, 1781–1850, and 1851–1920), which are the time-bands used in the CLMETEV corpus. As it turned out that the difference between the middle and the late periods were more often than not not significant, we collapsed the middle and the late period.

Finally, the following additional factors were distinguished:

(ix) *The type of subject of the main clause* (TYPE SUBJECT MAIN CLAUSE):

- pronoun1 = first-person pronoun
- pronoun2 = second-person pronoun
- pronoun3 = third-person pronoun
- noun
- none (when an explicit main clause subject is missing, as in *Remember his giving away every thing when he set out upon his grand expedition* (1794, CLMTEV, William Godwin, *The Adventures of Caleb Williams*))

- (x) ANIMACY SUBJECT CC. The subject may be:
- animate
 - inanimate
- (xi) VOICE VERB CC
- active
 - passive
 - copular
- (xii) DENOTATION. Two values are distinguished:
- ‘different’: main clause subject and CC subject denote different entities
(as in (16))
 - ‘same’: main clause subject and CC subject denote same entities; in
this condition, the CC subject may be controlled by the matrix subject
(as in (17a)) or not (as in (17b))
- (16) *Do you remember a gang coming there in Oct. 1746?* (OBC, t17500711–31)
- (17) a. *Debbenham denied ever knowing Mr. Athill.* (OBC, t17970712–71)
- b. *He denied his ever giving such an Order to the Prisoner.* (OBC,
t17280228–45)
- (xiii) TEMPORAL RELATION. Indicates the temporal relation between the CC and the
time of the CTP:
- anterior (18a)

- simultaneous (18b)
 - posterior (18c)
- (18) a. *I remember a detective coming and speaking to me.* (1881, OBC, trial id = t-18811212–125)
- b. *My lady and Miss Rachel regret that they are engaged, Colonel.* (1868, CLMTEV, William Collins, *The Moonstone*)
- c. *Remember that I shall see you in the summer.* (1746–71, CLMTEV, Philip Chesterfield, *Letters to his Son*)

(xiv) COMPLEMENTATION TYPE. Four types of CC are distinguished:

- *-ing* CC (includes present as well as perfect gerundials)
- *to*-infinitive CC (includes present as well as perfect infinitives)
- *that*-clause
- zero-complementizer clause (a *that*-clause without a *that*-complementizer)

10.5 A probabilistic analysis of complement choice

10.5.1 Charting the distribution of CCs in LModE

Table 10.1 presents the distribution of finite vs. non-finite CCs with the CTPs *remember*, *regret*, and *deny* across our two LModE sub-periods. It can be seen that the aggregate frequencies show an increase of non-finite vs. finite CCs: from a slightly higher share of non-finite CCs in the Early period (50.24% of all CCs) to a substantially higher share in the Late period (68.53%). For the individual verbs,

however, distributions vary: while there is an marked increase of non-finite CCs with *remember* (from a 49.12% share of non-finite CCs in the Early period to 76.01% Late period), it is finite CCs which outnumber non-finite CCs with *regret*, while *deny* shows stable variation.

@@ insert Table 10.1 here @@

10.5.2 What determines the choice of non-finite CCs?

It can be observed that none of the factors/values proposed uniquely conditions one outcome or the other (*that*-clause or non-finite clause). Consider in this respect (19) and (20). While the condition DENOTATION = ‘same’ triggers an *-ing*-CC in (19), it is also a condition that allows a *that*-CC, as in (20).

(19) *I remember on one occasion asking Judd if Mr. Vyse was in.* (OBC, t-18741123–39)

(20) *I remember when I was a boy myself I was once asked to take tea at a girls' school.* (1903, CLMETEV, Samuel Butler, *The Way of All Flesh*)

Similarly, the choice of an *-ing*-CC in (21) can be hypothesized to be attributed to its simple VP structure, consisting of a verb only, without an additional argument or modifier. However, attestations such as (22) show that the condition COMPLEXITY CC = ‘V’ is not sufficient to predict the *-ing*-CC.

(21) *I don't remember any other persons calling.* (OBC, t-18710227–190)

(22) *I do not remember he ran.* (OBC, t17700221–19)

Conversely, *that*-clauses can be hypothesized to be associated with the condition DENOTATION = ‘different’ (i.e., different denotation of matrix subject and CC-subject) and with the CC’s complex VP structures (condition COMPLEXITY CC = ‘Vobj+adv’), as exemplified in (23):

- (23) a. *I do not remember that he has made my payments before.* (OBC, t-18610408–350)
b. *I don’t remember that I saw Miles Barne at the fire.* (OBC, t17670115–32)

However, our data show that each of these conditions can also be observed in non-finite *-ing*-CCs, as in (24):

- (24) a. *I remember a detective coming and speaking to me.* (OBC, t-18811212–125))
b. *Do you remember asking the prisoner after breakfast to watch while you got to the bottom of the street ...* (OBC, t-18360404–919)

In other words, the factors identified above are predictors¹² of the non-finite as well as the finite CC-types following *remember/regret/deny*. In particular, each condition within a particular factor/predictor favours non-finite as well as finite CCs.

Importantly, these patterns of preference vary across factors; in other words, one

¹² In a statistical model (such as the one described below), the factors triggering a particular outcome are said to have a *predictive* value (see Gries 2012)

factor's relative weight or impact on the choice of finite vs. non-finite CC-type may differ from another. Compare Table 10.2 representing the distribution of finite vs. non-finite CCs for the factor INTERVENING MATERIAL IN WORDS with Table 10.3 representing this distribution for the factor MEANING CC. In Table 10.2, the condition INTERVENING MATERIAL = '0 words' has a preference for non-finites whereas the condition INTERVENING MATERIAL = '>1 words' has a preference for finites. In Table 10.3, non-finites are preferred over finites in both conditions, but there is a higher preference for non-finites in the condition MEANING CC = 'event/action'.

@@ insert Table 10.2 here @@

@@ insert Table 10.3 here @@

At first sight, then, it would appear that frequency distributions might be an indicator of differential impact of the factors, in other words of the relative weight of the conditioning factors. However, when comparing Table 10.3 (above) with Table 10.4, it can be seen that frequency distributions for two factors can also be fairly similar.

@@ insert Table 10.4 here @@

Indeed, Table 10.3 shows that in the condition MEANING CC = 'event/action', there is a higher probability of the non-finite CC being chosen than when the CC denotes a state (70% vs. 52% preference for non-finite). Table 10.4 shows a similar distribution in that the condition TEMPORAL RELATION = 'anterior' favours non-finite in 66% of the cases and TEMPORAL RELATION = 'simultaneous' in 55% of the cases. As such, it is not clear to what extent the higher proportion of non-finite vs. finite CCs is to be

attributed to the ‘event/action’ meaning of the CC or to the ‘anterior’ meaning of the CC; in other words, it is not clear what the relative strength of either condition is.

In light of this, it would appear that frequency distributions alone cannot sufficiently inform us about the relative strength of each of the variables/factors. In a situation such as this, where we are dealing with probabilistic complementation choice in a context where various factors are at play, it is advisable to fit a regression model.

To probe the multivariate and probabilistic nature of complementation strategy choice, we fit a *binary logistic regression model* with mixed effects (Pineiro and Bates 2000)¹³. Logistic regression modelling is the closest a corpus analyst can come to conducting a controlled experiment: the technique models the combined contribution of all the conditioning factors considered in the analysis, systematically testing the probabilistic effect of each factor while holding the other factors in the model constant. Ideally, the factors should be independent of each other, but in practice a certain degree of overlap (in technical parlance, *collinearity*) is admissible. The random effects in the model take care of non-central and not necessarily repeatable factors such as author idiosyncrasies, register variation, and differential propensities of different verb lemmas and meanings to co-occur with either complementation type.

Our response variable is binary, distinguishing between finite complementation (*that* + zero) and non-finite complementation (all other observations). Predicted odds are

¹³ We utilize the lme4 package in the statistical software package R (R Development Core Team 2011).

for non-finite complementation. As was pointed out earlier, right at the outset the ‘remember2’ and ‘remember3’ observations were excluded due to lack of variability. We observed the customary steps to obtain a minimal adequate regression model: We began by fitting the maximal model including all potentially important language-internal factors. Subsequently, the model was simplified by removing factors lacking significant explanatory power (for instance, we removed the predictor [CTP – CC] TIME-REFERENCE, because it did not turn out to have a significant effect). The resulting model is of acceptable quality.¹⁴

We begin by briefly addressing the three random effects in the model. MEDIUM (variance: 1.77, standard deviation: 1.33) accounts for most of the variance not accounted for by the fixed effects in the model. Specifically, speech-based texts favour non-finite complementation, while written texts favour finite complementation. As for verb lemmas and verb meanings (variance: 0.85, standard deviation: 0.92), we find that ‘remember1’ favours non-finite complementation, while the other verb lemmas and meanings distinguished in the present study (‘regret1’, ‘deny1’, and regret2’ – in that order) disfavour it. Finally speaker/writer

¹⁴ Somers D_{xy} , a rank correlation coefficient between predicted outcome probabilities and observed binary outcomes, is .81, which indicates that the model discriminates fairly between complementation types. The model correctly predicts 84% of all complementation outcomes in the dataset, which is a significant increase over baseline (64.1%) predicting the overall most frequent value. The model's condition number κ is 21.7, which is another way of saying that the model exhibits moderate though not harmful (Baayen 2008:182) collinearity.

idiosyncrasies – approximated here by modelling CORPUS FILE ID as a random effect – is only of minor importance (variance: 0.62, standard deviation: 0.79).

@@ insert here @@

Table 10.5 summarizes the fixed main effects in the model. The figures reported in the table are so-called odds ratios (henceforth: ORs), which quantify the magnitude and the direction of the effect of each factor on the odds that non-finite complementation is chosen. In the case of categorical predictors (such as MEANING CC), ORs measure an effect against the backdrop of an (arbitrarily chosen) default value; for continuous predictors (such as INTERVENING MATERIAL IN WORDS), ORs measure the effect of one-unit increases. Because odds ratios can take values between 0 and ∞ , two cases can be distinguished:

1. if $OR < 1$, the predictor makes the outcome ‘non-finite complementation’ less likely
2. if $OR > 1$, the predictor makes the outcome ‘non-finite complementation’ more likely.

Thus we find that MEANING CC has a robustly disfavoured effect on complementation choice: if the value is ‘state’ and not ‘event/action’, the odds for non-finite complementation decrease by a factor of 0.36, i.e., by 64%. INTERVENING MATERIAL IN WORDS also discourages non-finite complementation (for every additional word, the odds decrease by a factor of 0.73, i.e., by 27%), and so do complex NP subjects of complement clauses (TYPE SUBJECT CC): if the subject is complex, the odds for non-

finite complementation decrease by a factor of 0.56. Also disfavoured is (i) the DENOTATION = 'different' condition (OR: 0.18), (ii) an inanimate CC subject (CC SUBJECT ANIMACY) (OR: 0.47), (iii) a TEMPORAL RELATION of anteriority (as opposed to simultaneity) between matrix and CC clause (OR: 0.37), and (iv) the early PERIOD (OR: 0.50), which spans the years between 1710 and 1780. This is another way of saying that non-finite complementation is more common after 1780 than before 1780 (and that the other main effects sketched in Table 10.5 describe the late period, thanks to the default level of PERIOD being set to 'late').

Which factors favour non-finite complementation? For one thing, vis-à-vis TYPE SUBJECT MAIN CLAUSE = 'none', all other categories of the factor favour non-finite complementation – most so 'pronoun2', which increases the odds for non-finite complementation by a factor of about 11, and least so 'noun', which increases the odds for non-finite complementation by a factor of 'only' 3.57. Second, consider COMPLEXITY CC: in comparison to CC verbs without argument/modifier, both verbs with one argument/modifier ('Vobjadv') and verbs with an argument + modifier, or two arguments, or two modifiers ('Vobj+adv') favour non-finite complementation with ORs of about 2. Third, vis-à-vis VOICE OF CC VERB = 'active', VOICE OF CC VERB = 'passive' increases the odds for non-finite complementation by a factor of 2.15.

Finally, we turn to the issue of how the probabilistic effect of various conditioning factors has changed during the LModE period. To this end, Table 10.6 reports significant interaction terms between REAL TIME PERIOD and three language-internal constraints. (Note that the figures reported in the table are factors by which we have to multiply the main effects in Table 10.5 to gauge their effect in the early period.)

@@ insert Table 10.6 here @@

First, observe that according to Table 10.5, TYPE SUBJECT CC = ‘complex NP’ disfavours non-finite complementation at 0.56. What we learn from Table 10.6 is that in the early period, TYPE SUBJECT CC = ‘complex NP’ actually favoured non-finite complementation, at $0.56 * 4.01 = 2.25$. Put differently, TYPE SUBJECT CC has reversed its effect in the course of the LModE period. Second, DENOTATION = ‘different’ disfavours non-finite complementation throughout, but more so in the late sub-period than in the early sub-period. Thus, the DENOTATION constraint is subject to a diachronic *strengthening* in LModE. Third, TEMPORAL RELATION = ‘anterior’ disfavours non-finite complementation throughout, but more so in the early period than in the late period. The TEMPORAL RELATION constraint is hence subject to a diachronic *weakening* in LModE.

The main results from our regression analysis can be summed up as follows:

- The only semantic factor significantly predicting non-finite vs finite CCs is whether the CC expresses a state or an event;
- structural factors play an important role in determining CC-type: the factors INTERVENING MATERIAL and TYPE SUBJECT CC each significantly disfavour non-finite complementation;
- the speech-based genre is more favourable to non-finite complementation;
- over time, there appears to be a significantly higher preference for non-finite CCs.

An additional interesting result is that the condition DENOTATION = ‘different’ disfavors non-finite complementation. This is not likely to be surprising because in 72.5% of the cases (1845/2548) that meet the DENOTATION = ‘same’ condition, the CC-subject is controlled by the matrix subject, i.e., is left unexpressed, and as such requires a non-finite complement. In line with these findings, the condition CC SUBJECT ANIMACY = ‘inanimate’ also disfavors non-finite complementation; indeed, inanimate CC subjects, on the whole, differ from the matrix subject, which is a condition disfavoring non-finite complementation.

In our regression model, the factor [CTP – CC] TIME REFERENCE did not turn out to have a significant effect on CC-choice. This does not mean, however, that the frequency distributions associated with this factor are uninformative (see section 10.6).

10.6 Interpretation and discussion

In this final section, we explore how the results of this study bear on earlier hypotheses about CC-variation. Let us first consider how the factors relating to structural complexity impact Rohdenburg’s Complexity Principle.

In various publications, Rohdenburg (1995, 1996) claims that CC-variation can be accounted for in terms of the ‘Cognitive Complexity’ principle. This principle states that in a cognitively more complex environment, more explicit grammatical options tend to be favoured (see Rohdenburg 1996: 151). A cognitively more complex environment, in turn, is hypothesized by Rohdenburg as requiring more processing

effort. Grammatical manifestations of cognitive complexity are, for instance, constituents of sizeable length and great distance between head and dependent (notably because of insertions), in other words, structurally complex entities or environments. If we take structural complexity as a measure of cognitive complexity, it follows that these structurally complex entities/environments will tend to favour the more explicit finite *that*-CCs (and less complex environments the less explicit non-finite CCs, namely gerundial *-ing*-clauses and *to*-infinitives) – *that*-CCs can be seen as more explicit grammatically because they carry tense/ mood distinctions and obligatorily have an expressed subject.

Our analysis has shown that the factors INTERVENING MATERIAL and TYPE SUBJECT CC, each associated with structurally complex environments, significantly disfavour non-finite complementation. As such, our results strengthen Rohdenburg's earlier findings: they are not only consonant with his observations (with respect to the variation *to*-infinitive vs. *that*-clause) that INTERVENING MATERIAL has a disfavoured effect on non-finite complementation, but they reveal additional complexity factors having a similar disfavoured effect. At the same time, however, it has been seen that complex CC predicates (COMPLEXITY CC) in fact favour non-finite CCs (and to the extent that passive verbs can be considered to be more complex than active verbs, a similar favouring effect can be observed). In other words, Rohdenburg's proposed disfavoured effect cannot be generalized to all structural complexity factors, and in that sense, Rohdenburg's Complexity Principle does not apply as generally as commonly held.

Second, Rohdenburg (1995) finds that in the course of the seventeenth and eighteenth centuries, *that*-clauses are increasingly replaced by *to*-infinitives. In a recent study on the evolutionary path of *that*-clauses vs. gerundial *-ing*-clauses, however, Rohdenburg (Chapter 9, this volume) sketches a more qualified picture, in that any replacement of *that*-clauses by gerundial *-ing*-clauses varies with individual matrix verbs. Now, even though the focus of our study is on shifting/changing patterns of preference rather than on the phasing out of *that*-clauses (as in Rohdenburg's studies), our results are largely in line with Rohdenburg's observations.¹⁵ Indeed, developments for the individual verbs may vary (as can be inferred from the effects of the random factor VERB MEANING, i.e., differential propensities of the different verb meanings with either CC-type), but aggregate figures for the matrix verbs show an increase of non-finite CCs vs. finite CCs (see section 10.5).

The question, then, is if the results of our investigation can be tied up with results from grammaticalization research. If indeed, there is an increase of non-finite complement clauses relative to *that*-CCs, the question is whether this path of development – from a relatively loose concatenation of [matrix verb + *that*-clause] to a *to*-infinitive or *-ing* CC showing more unified/bonded clause combining – can be considered as a case of grammaticalization?

The question is justified in the light of Lehmann's (1995) characterization of grammaticalizing structures as losing autonomy. Further, it seems to tie in with Lehmann's (1988) and Hopper & Traugott's (2003) suggestion that clause linkage

¹⁵ Denison has made a similar claim ('A long-term trend in English has been the growth of nonfinite complement clauses at the expense of finite clauses' (1998: 256); this claim is, however, not based on extensive corpus study).

and in particular the development of complementation patterns can also be seen as a grammaticalization process, in that verb complements may be subject to ‘deategorialization’ (in a way similar to the decategorialization of lexical items) when being integrated with a main clause. As such, verb complements may show a path of development “from a loose, paratactic concatenation via syntacticization into non-finite embedding” (Givón 1979: 214), whereby the non-finite shows a reduction in tense-aspect morphology and lack of subject agreement, may (at least for English) have become reduced to a *to*-infinitive or gerundive *-ing* clause, and may thus show more unified/bonded clause combining (see Hopper & Traugott 2003: Chapter 7).

While an increase, over time, in the percentage of non-finite CC tokens over *that*-CC tokens always represents a path of gradually stronger integration of the complement clause into the matrix, it is at first sight doubtful that this shift from a finite (looser) *that*-CC to a more bonded non-finite CC should be seen as a case of grammaticalization. For one, the non-finite and finite structures have different origins; as such, one might want to talk of replacement of the finite CC by a non-finite CC rather than of grammaticalization through decategorialization (see Fischer 2007: 221–222). Further, a decategorialized, that is, reduced, clause is still as much a grammatical structure as a full clause; in other words, decategorialization here does not correlate with loss of lexical characteristics.

What may, to some extent, salvage the idea of grammaticalization is that, as the finite clauses are gradually being replaced by non-finite clauses, as can be seen with ‘remember1’ (see Table 10.1), the non-finite becomes gradually more obligatory. In other words, the choice between the members of a paradigm becomes more

constrained, in that once the speaker has decided on a particular matrix verb with a particular meaning, the choice of the CC becomes more obligatory (for a similar view, see Haspelmath 1989: 298; see also Lehmann 1995 on the loss of paradigmatic variability).¹⁶ Additional cases of loss of paradigmatic variability can be observed in the alternation between non-finites whose subject is controlled by the matrix subject (and hence, whose subject is the same as the matrix subject and remains unexpressed) and non-finites whose subject is the same as that of the matrix subject (but is expressed). An example is (25):

- (25) a. *I remember my going into the House.* (OBC, t17370114–1)
b. *Do you remember at any time going to the prisoner's house?* (OBC, t17730217–52)

¹⁶ In fact, the use of notion of ‘obligatoriness’ is somewhat surprising, even seems contradictory, in a study of non-categorical variation. However, to the extent that the preference for non-finites with a particular matrix verb becomes more outspoken over time, one can talk about ‘increasing obligatoriness’, and hence ‘increasing grammaticalization’. So, in our study of non-categorical variation, there seems to be *some* categorical distribution of variants that would provide an argument in favour of grammaticalization, but other instances that argue against it. Rather than treat this issue as verb meanings *grammaticalizing* with a particular CC-type, one might more successfully deal with it as *constructionalization* (see, for instance, Trousdale 2012, Traugott and Trousdale 2013) (thanks are due to Marianne Hundt for raising this point).

Table 10.7 shows that the relative proportion of non-finites with an expressed subject that is identical with that of the matrix becomes increasingly smaller: in other words, these types of expressed subject are gradually ousted from the paradigm of CC subjects.

In sum, the analysis of change and variation with *remember*, *regret*, and *deny* shows that the grammaticalization from lexical items/constructions cannot simply be transferred to processes involving sentential/clausal units (in particular to verb complementation and clause linkage). Evidently, further analysis of several sets of matrix verbs is needed to reach more definitive conclusions. As such, the results found in this study seem at least to tie in with Fischer's uneasiness about the (alleged) grammaticalization path from *that*-clauses to non-finite *to*-infinitives with verbs of volition (see Fischer 2007: 221).

Finally, we will explore whether our observed preferences in CC-choice can be usefully linked up with Noonan's hypothesis about reduced complement clauses (1985: 100 –101). Noonan (1985: 100 – 101) points out that "reduced complements, which are likely to lack tense distinctions ... are typically associated with predicates whose complements have DTR [dependent time reference]." As such, he puts forward that "infinitives ... are frequently restricted to DTR contexts ... [and that] indicative complements are normally excluded from DTR contexts since they are typically coded for tense, and therefore the expression of tense in such cases is redundant." A similar view is expressed by De Smet (2012b: 142). Noonan's and De Smet's views are supported by our study. While the factor CTP-CC TIME REFERENCE did not have a significant effect on CC-choice, the frequency distribution associated with this factor

may be instructive. The matrix verb ‘remember₂’ = ‘not to forget to do something’ (which was left out of the analysis because it does not show CC-variation), specifies a temporal domain in the future with respect to the time of the matrix verb, and categorically takes a *to*-infinitive.

- (26) *Remember to make a great difference between companions and friends.* (1746, CLMETEV, Philip Chesterfield, *Letters to his Son*)

Then again, the matrix verb ‘remember₃’ = ‘keep in mind’ (which was left out of the analysis as well), combines exclusively with a finite *that*-clause and has ITR complements (i.e., complements may refer to anterior, simultaneous, or posterior time domains with respect to the time of the matrix).

- (27) a. ‘anterior’: *Remember, I’ve known and liked him steadily for nearly three years.* (1910, CLMETEV, Edward Forster, *Howard’s End*)
- b. ‘simultaneous’: *But you must remember Aldous is no boy.* (1894, CLMETEV, Humphrey Ward, *Marcella* (Vol. 1))
- c. ‘posterior’: *Remember that I shall see you at Hanover next summer.* (1746, CLMETEV, Philip Chesterfield, *Letters to his Son*)

Of course, our data do not always show a 1:1 relation between DTR and non-finite or between ITR and finite. However, when the semantics of the matrix verb specifies the temporal and modal status of the event in the complement clause, there is at least a higher proportion of non-finite CCs. This is the case with ‘remember₁’ = ‘recall’, which typically specifies a past event/situation, and can therefore often do without a

full tense specification in the CC; the aggregate figures for ‘remember’ (which is largely identical to ‘remember1’) attest to this. Conversely, if the matrix verb is neutral with respect to the CC’s temporal and modal status, there is a higher proportion of finite CCs (compared to the share of finite CCs with CTPs such as *remember*). This situation holds for *deny*, which does not specify any particular temporal domain, and whose share of finite vs. non-finite CCs is 45% and 55%, respectively.

10.7 Conclusions

In this chapter, we examined the probabilistic or non-categorical variation in LModE of finite vs. non-finite complement clauses with the matrix verbs *remember*, *regret*, and *deny*. To probe the multivariate nature of complementation strategy choice, we fit a *binary logistic regression model* with fixed effects to the data, thus shedding light on the factors favouring non-finite complementation and on the issue of how the probabilistic effect of various conditioning factors changed during the LModE period. Finally, we considered the relevance of the observed preferences (i) for Rohdenburg’s (1996) complexity hypothesis, (ii) for Denison’s (1998) and Rohdenburg’s (1995, this volume) views on the replacement of finite *that*-clauses by non-finite clauses, and concomitantly for our position on the grammaticalization of sentential/clausal units, and (iii) for Noonan’s (1985) views on reduced clauses.

Our study shows that Rohdenburg’s complexity principle needs to be qualified, in that not all types of structural complexity have a discouraging effect on non-finite complementation. While developments for the individual verbs may vary, aggregate figures for the matrix verbs showed an increase of non-finite CCs vs. finite CCs, thus

largely corroborating observations by Denison (1998: 256) and Rohdenburg (1995, this volume) that there has been a growth of nonfinite complement clauses at the expense of finite clauses. Finally, support was also found for Noonan's claims on reduced clauses.

While our analysis is, on the whole, not at odds with earlier claims on the differential patterns of preference and differential evolution of finite and non-finite CCs, it is important that it has made use of state-of-the-art methodology – to our knowledge, this is the first study applying statistical modelling to the CC-variation – and involves a large data set, thus putting these earlier analyses on much more solid empirical ground. Further, while the focus in this chapter has been on non-categorical CC-variation, and on statistical modelling as a way to get a handle on this variation, a comprehensive account of CC choice in LModE also needs to give due attention to categorical CC-variation (which cannot be examined in terms of probabilistic modelling) – in fact, the two kinds of study complement each other. Finally, it is also good to keep in mind frequency distributions still have a role to play in a study of CC-variation.