The Complementation of the Adjective *Sorry* in Recent Centuries

Nasti Tikkanen

University of Tampere

School of Language, Translation and Literary Studies

English Philology

Pro Gradu Thesis

April 2015

Tampereen yliopisto
Englantilainen filologia
Kieli-, käännös- ja kirjallisuustieteiden yksikkö

TIKKANEN, NASTI: The Complementation of the Adjective *Sorry* in Recent Centuries

---

Tämä pro gradu –tutkielma käsittelee adjektiivin *sorry* komplementaatiota britannianenglannissa vuodesta 1710 nykypäivään. Tutkielman päätavoitteena on selvittää, miten adjektiivin komplementaatio on muuttunut vuosisatojen kuluessa ja millaisia komplementteja sen kanssa käytetään nykykielessä.

Tutkimusaineistona käytetään korpusesimerkkejä kahdesta eri korpuksesta. *The Corpus of Late Modern English Texts* (CLMET) kattaa vuodet 1710–1920 ja nykyenglannin aineisto (1960–1993) on saatu *The British National Corpus* (BNC) –nimisestä korpuksesta. CLMET on edelleen jaettu kolmeen samankokoiseen osaan: CLMET 1 sisältää tekstejä vuosilta 1710–1780, CLMET 2 keskittyy vuosiin 1780–1850 ja CLMET 3:n tekstit on julkaistu vuosina 1850–1920. CLMET:n aineisto koostuu kaunokirjallisista teoksista, mutta BNC sisältää monia eri tekstilajeja. Tämän takia BNC:n materiaalista otettiin huomioon vain kaunokirjalliset tekstit, jotta vertailtavat esimerkit olisivat mahdollisimman samankaltaisia.

Tutkielman ensimmäisessä osassa tarkastellaan komplementaatioon liittyviä teorioita ja korpuslingvistiikkaa sekä esitellään tutkimuksen aineistona toimivat korpukset. Lisäksi tutustutaan *sorry*-adjektiivin ominaisuuksiin englanninkielisten sanakirjojen ja kielioppiteosten avulla. Pääasiallisena sanakirjalähteenä käytetään tunnettua *The Oxford English Dictionarya* (*OED*) ja sitä täydentämässä on kaksi edistyneille kielenopiskelijoille suunnattua sanakirjaa, *Oxford Advanced Learner's Dictionary* sekä *Collins Cobuild Advanced Learner's Dictionary*. Tutkielman jälkimmäisessä osassa tarkastellaan ja analysoidaan adjektiivin komplementaatiota sekä siinä tapahtuneita muutoksia kronologisessa järjestyksessä neljän korpusosan (CLMET 1–3 sekä BNC) mukaisesti.

Tutkimus osoittaa, että vuosina 1710–1850 *sorry*-adjektiivin yleisin komplementti oli *to*-infinitiivi, mutta vuodesta 1850 eteenpäin *sorry* on useimmiten esiintynyt ilman komplementtia. *To*-infinitiivi ei ole kuitenkaan täysin kadonnut, ja muitakin vaihtoehtoja kuten *that*-lauseketta ja *for* + NP –komplementtia käytetään yhä. Esimerkiksi *about*-preposition käyttäminen *sorry*-adjektiivin kanssa on itse asiassa hieman yleistynyt vuoden 1920 jälkeen. Sen sijaan CLMET 2:ssa havaitut komplementit *at* + NP sekä *at* + –*ing*-muoto eivät enää esiinny nykykielessä. Komplementtien kokonaismäärä on kuitenkin pysynyt suurin piirtein samana. Tutkimuksessa tuli esiin myös jatkuvasti yleistyvä taipumus käyttää adjektiivia ilman subjektia ja verbiä – komplementista riippumatta.

Asiasanat: sorry, komplementaatio, korpus, korpuslingvistiikka, adjektiivi

Table of contents

1 Introduction

Consider the following examples of the adjective *sorry* in various different constructions:

(1) I am very *sorry* that it will be impossible for me to see you next Sunday. (Gissing, *The Odd Women*, 1893)

(2) 'Oh, I am *sorry* to hear that,' said Knight, in a changed tone. (Hardy, *A Pair of Blue Eyes*, 1873)

(3) "Ah!" said Roger again; "I'm *sorry* for James; he had trouble with Dartie." (Galsworthy, *The Man of Property*, 1906)

(4) "Henry, I am awfully *sorry*." (Forster, *Howards End*, 1910)

These examples, all taken from *The Corpus of Late Modern English Texts*, illustrate the versatility in the use and complementation of *sorry*. Its possible complements include, for example, *that*-clauses (1), *to*-infinitives (2), and prepositional phrases (3). However, *sorry* can also occur on its own without any complements (4).

The aim of this thesis is to examine the complementation patterns of the adjective *sorry* in recent centuries with the help of two different corpora containing authentic examples of written English: the *Corpus of Late Modern English Texts* (CLMET), which includes works published between the years 1710 and 1920, and the more recent *British National Corpus* (BNC), which provides us with material from 1960 to 1993. The fact that the data in these corpora has been collected from three different centuries allows me to compare the possible changes in the use and complementation of *sorry* in different time periods, which may be of interest in understanding the behaviour of the word in question. Some of the complements we, as speakers of Present-Day English, use today may be relatively recent additions to English usage, whereas other complementation patterns may have been thriving at some point but are now almost completely forgotten. My intention is to find out whether there is any evidence of that kind of development in the recent history of *sorry*.

Another interesting viewpoint concerns the meanings of the patterns. The word *sorry* can

mean different things in different environments: in example (3) above, for instance, it means that rather than apologising for something, the speaker actually feels sympathetic towards James. In this thesis, I want to examine the relationship between form and meaning and also see if there are signs of any patterns changing their meaning or the way they are used during the last three centuries.

The first part of the thesis (chapters 2, 3 and 4) is the introductory part in which I will explain the important terms, concepts and methods associated with complementation and corpora, and have a look at what has been said about this subject in previous literature. The second half, on the other hand, will focus on my own analysis of the corpus data and subsequent discussion of the findings.

2 On complementation

In this chapter, I will explain what is meant by the term 'complementation' as well as explore some key aspects that have an influence on complement selection and may therefore affect the complementation of *sorry*.

2.1 What is a complement?

Quirk et al. define complementation as "the function of a part of a phrase or clause which follows a word, and completes the specification of a meaning relationship which that word implies" (1985, 65). In other words, complements are words or groups of words that are needed to complete a meaning that is conveyed by a certain grammatical construction (Leech and Svartvik 2002, 271). Huddleston and Pullum, on the other hand, describe complements by stating that their "most important property … is that they require the presence of an appropriate verb that licenses them" (2002, 219). Not all verbs are appropriate in any given construction, because different verbs license different complements. Consider the following examples (ibid.):

> (1a) She <u>mentioned</u> the letter.
> (1b) *She <u>alluded</u> the letter.
>
> (2a) She <u>thought</u> him unreliable.
> (2b) *She <u>said</u> him unreliable.

Huddleston and Pullum refer to this relationship between verbs and complements as *subcategorisation*: verbs are subcategorised according to the complements they choose. In the example above, *mentioned* and *alluded* are not interchangeable, because they belong to different subcategories of verbs and therefore do not accept the same complement constructions.

However, not all the parts of a phrase or clause that follow a specific word can be considered as its complements. Compare the following examples from Somers (1984, 508):

(3) He looked for his friend *in London*.
(4) James lives *in London*.

These superficially similar-looking constructions are actually very different. The italicized part in (3) is an adjunct, but the exact same prepositional phrase, consisting of a preposition and a noun phrase, has the function of a complement in (4). This can be easily seen when the prepositional phrase is omitted: *He looked for his friend* makes perfect sense, but *James lives seems to be lacking some information that is essential for the full understanding of the utterance's message and meaning. Somers clarifies this distinction by noting that adjuncts are "essentially optional elements which can be said to complete the meaning of the central predication as a whole", whereas complements are usually obligatory or at least "closely associated with the verb" (ibid.). This difference between adjuncts and complements will be discussed further in the following section.

The examples in this section have focused entirely on verbs and their complements. However, at this stage it must be pointed out that adjectives can take complements as well – as the topic of this thesis suggests. Leech and Svartvik (2002, 227) list prepositional phrases, *that*-clauses and *to*-infinitives among possible complements of adjectives and provide the following examples:

(5) I feel very sorry *for Ann.*
(6) Everybody's pleased *that she is making such good progress.*
(7) I'm glad *to hear she is recovering.*

2.2 The obligatoriness of complements and valency theory

Huddleston and Pullum (2002, 221) point out that adjuncts are always optional, but that there are both optional and obligatory complements. They illustrate this distinction with the following examples (ibid.):

(8a) She left <u>because she was ill</u>.
(8b) She left.

(9a) She read <u>the report</u>.
(9b) She read.

(10a) She perused <u>the report</u>.
(10b) *She perused.

In the examples above, the underlined part is an adjunct in (8a), an optional complement in (9a) and an obligatory complement in (10a). As was mentioned above, obligatory complements are quite easy to distinguish from adjuncts (and optional complements) by trying to omit them. The phrase will become ungrammatical if the omitted item is an obligatory complement (example 10b). As for the difference between adjuncts and optional complements, Huddleston and Pullum state that "if an element is obligatory, and hence a complement, with some verbs, then in the absence of counter-evidence we will take it to be a complement rather than an adjunct when it is optional too" (ibid.). They point out that although the complement is optional in (9a) and obligatory in (10a), the examples are so similar in other respects (e.g. passivization: *The report was perused/read by her*) that they cannot be said to represent entirely different constructions. Therefore, *the report* must be a complement in both examples, even though the degree of obligatoriness is different.

But how do we distinguish adjuncts from optional complements? According to Huddleston and Pullum, the choice of position in the sentence is more restricted for complements (ibid., 225). For example, the phrase *We played tennis in the afternoon* can be paraphrased as *In the afternoon, we played tennis* and still be perfectly acceptable, but *To Kim he gave the beer* is not as acceptable as *He gave the beer to Kim.* Therefore, it seems that adjuncts can be moved around more freely.

Another useful test for distinguishing complements from adjuncts is the so-called *do so* test. According to Lakoff and Ross (1966, II-4), *do so* is "a pro-form which may substitute for a verb

phrase". Therefore, elements which can occur after it without making the sentence ungrammatical are outside of the verb phrase. Conversely, elements which cannot follow *do so* are included in the verb phrase – i.e. complements of the verb (ibid., II-5). The *do so* test can be further illustrated with the help of the following examples (Huddleston and Pullum 2002, 223):

> (11a) *Jill keeps her car in the garage but Pam does so on the street.
> (11b) Jill washes her car in the garage but Pam does so on the street.

Huddleston and Pullum explain that the "antecedent for *do so* has to embrace all the internal complements of the verb: it therefore cannot itself combine with such a complement." (ibid.) This means that in (11a) *in the garage* is an obligatory complement, because in this construction, *Pam does so* actually means the same as *Pam keeps her car in the garage.* Therefore, *Pam does so* cannot be followed by *on the street* without making the sentence ungrammatical: *does so* already includes one complement. However, in (11b) *Pam does so* only includes *Pam washes her car*, which means that in this case, *in the garage* is an adjunct and *does so* can be followed by a complement.

Valency theory is an approach which emphasises this distinction between adjuncts and complements. According to Herbst et al (2004, xxiv), valency theory considers the verb to be in a central position in a sentence, because it "determines how many other elements have to occur in order to form a grammatical sentence" (ibid.). These "elements" are also known as complements, and the valency of a verb determines how many complements it takes. Some verbs only require one other element, while others need several elements to form sensible utterances. Because adjuncts are optional and therefore not essential in order for the verb to form a grammatical construction, they are not dependent on the verb and their form is not governed by it.

2.3 Argument structure and thematic roles

Although closely related to valency theory, argument structure and theta theory are concepts which take a slightly different approach to the relationship between the head and its complements. Argument structure focuses on the number and type of arguments a predicate can take in order to function grammatically. According to Haegeman (1991, 36), "arguments are the participants minimally involved in the activity or state expressed by the predicate". Consider the following examples (ibid., 38-40):

(12) Poirot is restless.
(13) Poirot will analyse the data.

The adjective in (12), *restless*, is a so-called one-place predicate, as it only requires one argument (in this case, *Poirot*) in order to function. *Analyse* in (13), however, is a two-place predicate: it needs both *Poirot* and *the data* to make sense. Omitting the second argument would result in **Poirot will analyse*, which tells us nothing about what is being analysed, and therefore does not make much sense. Similarly, **will analyse the data* seems to be missing a key ingredient – that is, the *agent.*

This brings us to the concept of thematic relations, which Carnie defines as "particular semantic terms that are used to describe the role that the argument plays with respect to the predicate" (Carnie 2007, 221). Even though the overall number and indeed the very nature of thematic roles are still under discussion (e.g. Dowty 1991, 547), there are a few basic roles which most linguists can agree upon. They are presented in the following table (adapted from Haegeman and Carnie):

Thematic role	Definition	Example
Agent	The initiator or doer of an action	*Brad* hit Andrew.
Experiencer	The entity that experiences some (psychological) state expressed by the predicate	*Keziah* likes cookies. *Becki* saw the eclipse.
Theme	The entity that undergoes actions, is moved, experienced or perceived	Shelley kept *her syntax book*. The arrow hit *Michael*.
Benefactive/Beneficiary	The entity that benefits from the action expressed by the predicate	He bought these flowers for *Jason*.
Goal	The entity towards which motion takes place	Millie went to *Chicago*.
Recipient	A goal where there is change of possession	Julie gave *Jessica* the book.
Source	Entity from which motion takes place	Stacy came directly *from sociolinguistics class*.
Location	The place where the action occurs	We're all *at school*.
Instrument	The object with which the action is performed	Patrick hacked the computer apart *with an axe*.

Table 1: Thematic roles and their definitions adapted from Haegeman and Carnie.

Thus, in example (14) below, *Constance* has the role of agent, *the ball* is the theme and *towards Poirot* represents the goal.

(14) Constance rolled the ball towards Poirot.

On the other hand, in a similar-looking sentence such as *Galahad gave the detective story to Jane*, the prepositional phrase *to Jane* could be assigned the thematic role of either goal or beneficiary, which goes to show that the identification of thematic roles is not always easy (Haegeman 1991, 42). Nevertheless, the essential idea of the theta theory (or *Θ*-theory) is condensed into the following principle known as the *Theta Criterion* (Carnie 2007, 225):

a) Each argument is assigned one and only one theta role.
b) Each theta role is assigned to one and only one argument.

In the above-mentioned example, therefore, *to Jane* can be either the goal or the beneficiary, but it

cannot be both at the same time, as the Theta Criterion states that each argument is assigned one and only one theta role.

2.4 Control and raising

Now that we are familiar with theta roles, we can move on to discuss the principles of control and raising. Consider the following examples from Carnie (2007, 395):

(15) Jean is reluctant to leave.
(16) Jean is likely to leave.

While these two examples may seem quite similar on the surface, they actually represent two different structures: (15) is an example of control, and (16) exemplifies a raising construction. Both structures will now be discussed further in their own respective sections.

2.4.1 Control

Let's consider example (15) above again, this time in terms of theta roles. Here, the subject NP *Jean* must have the role of experiencer, as the predicate *is reluctant* clearly refers to *Jean*. However, there is another predicate in the sentence, *to leave*, which also seems to refer to *Jean*, but now as an agent. Keeping the earlier introduced Theta Criterion in mind, it is obvious that *Jean* cannot be both experiencer and agent at the same time. This is why we need the help of PRO (for 'null pronoun', ibid.), a "covert" or "empty" pronoun subject (Radford 1988, 313). Written out, the deep structure of example (15) above would look like this:

(15') Jean is reluctant [PRO to leave].

The problem of multiple theta roles is now solved: here, the argument of *is reluctant* is *Jean* (experiencer), and the argument of *to leave* is PRO (agent) – which, in fact, refers to the same subject, *Jean*. This kind of construction is called *subject control*, because PRO is interpreted to refer back to the matrix subject (Radford 1988, 324). However, there is also another type of control,

object control, in which it is the object of the main clause that controls what PRO refers to, as in the following example (Carnie 2007, 396):

(17) Jean persuaded Brian [PRO to leave].

One question remains: Why do we need PRO? Carnie admits that PRO is, in fact, "a null element to account for an apparent hole in our theory" (ibid., 405), because without it, fulfilling the conditions of the Theta Criterion would be impossible in cases such as (15) above. However, before a better theory comes along, PRO is a handy tool which is able to account for most of the data.

2.4.2 Raising

Raising (also known as *NP movement*) differs from control in the way that PRO is not needed in a raising construction. Example (16) above could be paraphrased as follows:

(16') It is likely that Jean will leave.

This reveals that *Jean* is actually not the original subject of the matrix verb: the subject of the predicate *be likely* is, in fact, the semantically empty *it*, which is referred to as an *expletive* or *pleonastic pronoun* (Carnie 2007, 228). Therefore, in the original example (16), the subject NP *Jean* has been "raised" from the lower, embedded clause (*that Jean will leave*) to matrix subject position. Postal (1974, 33) offers a rather neat description of this phenomenon: in raising constructions, "main clause subject NPs are actually former complement subjects which achieved their final locus through Raising". *Likely* is only one example of so-called "Raising Predicates" (Radford 1988, 443) which are somewhat easy to recognise due to the fact that they "permit a non-referential Subject like existential *there*" (ibid.) – or indeed the expletive *it*, as in example (16') above. Another example of a raising predicate would be the fairly common *seem*. Thus, the difference between raising and control constructions can be further illustrated with the following examples:

(18a) Jean is likely to leave.
(18b) It is likely that Jean will leave.

(19a) Jean is reluctant to leave.
(19b) *It is reluctant that Jean will leave.

Reluctant is not a raising predicate and therefore does not allow an expletive in subject position.

This is the case with the topic of this thesis, *sorry*, too: *Jean is sorry to hear that* is fine whereas **It*

is sorry that Jean hears that is, obviously, nonsensical. Therefore, *sorry* always occurs in a subject

control construction rather than a raising construction.

2.5 Factors bearing on complementation

In this section, I will introduce a few principles which have been proven to affect the choice of

complementation pattern in order to make it easier to recognise them later in the thesis when the

data is being analysed.

2.5.1 The complexity principle

The complexity principle was first introduced by Günter Rohdenburg in 1996. It states that "in the

case of more or less explicit grammatical options the more explicit one(s) will tend to be favoured

in cognitively more complex environments" (Rohdenburg 1996, 151). According to Rohdenburg

(ibid. 173), cognitively complex environments may include

1. discontinuous constructions of various kinds (for example, intervening adverbial phrases between the verb and its complement)
2. (the presence of) more or less complex surface objects preceding finite and nonfinite clauses
3. heavy subject expressions (including subordinate clauses)
4. complex subordinate clauses
5. passive constructions

More explicit grammatical options, on the other hand, are defined as constructions that are, for example, "bulkier" or thought to be more formal than their more compact counterparts (ibid., 152). The bulkiness can manifest itself as an optional grammatical signal, as can be seen in the example below (Rohdenburg 1996, 151-2) in which the infinitive marker *to* makes the construction bulkier:

> (20) i I helped him to write the paper.
>   ii I helped him write the paper.

Alternatively, a choice between two function words can affect the bulkiness of the construction. The rivalling words in question can be morphologically related, but it is not necessary. This is, again, exemplified by Rohdenburg (ibid.):

> (21) She was prevailed on/upon to write another letter.

In this example, the preposition *upon* is thought to be the bulkier and more formal option.

Vosberg (2003b, 210) has also contributed to the topic of complexity by adding that insertions (material inserted between the matrix and lower clauses) in particular create above-mentioned discontinuous structures and therefore add to cognitive complexity. Thus, if the construction contains this kind of intervening material, we tend to favour more explicit grammatical options.

2.5.2 Extractions

Another factor which creates a cognitively complex environment is extraction which involves material that has been moved from the lower clause towards the left across the sentence boundary. Postal (1994, 162) lists nine different types of extractions:

> 1) Question extraction   Who$_1$ did they nominate t$_1$ to be director?
> 2) Restrictive relative extraction   The gun (which)$_2$ they claimed t$_2$ was used in the crime.

3) Pseudo clefting	What$_3$ Ellen wants t$_3$ is a Mercedes-Benz.
4) Negative NP extraction	[No such gorilla]$_4$ did I ever see t$_4$.
5) Comparative extraction	Stella tickled more chimps than (what$_5$) I said that Dwight tickled t$_5$.
6) Exclamatory extraction	[What a lovely woman]$_6$ I found out that he married t$_6$!
7) Topicalisation	Frank$_7$ I would never hire t$_7$.
8) Nonrestrictive relative extraction	Frank$_8$, who$_8$ they adored t$_8$, is dishonest.
9) Clefting	It was Frank who$_9$ they hired t$_9$.

In the above illustrations, also by Postal (ibid., 159), the extracted material is marked with a subscript and the *t* (for *trace*) represents the gap left behind by the extraction process. Vosberg (2003a, 307), on the other hand, lists relative extraction, comparative extraction, topicalisation and interrogation as the most important extraction types. He has also formulated the Extraction principle, which runs as follows (ibid., 308):

> In the case of infinitival or gerundial complement options, the infinitive will tend to be favoured in environments where a complement of the subordinate clause is extracted (by topicalisation, relativisation, comparativisation, or interrogation etc.) from its original position and crosses clause boundaries.

In other words, the *to*-infinitive complement tends to be favoured over the *–ing*-form in environments where extraction is present, as can be seen in the following example of relativisation (ibid., 307):

> (22) It is the worthy Spencer$_1$ whom I'm sure you remember [to have often heard [me mention t$_1$ in the relation of my private misfortunes]] … (John Dauncey, *The English Lovers*, 1622)

2.5.3 The *horror aequi* principle

Rohdenburg (2003, 236) defines the *horror aequi* principle as "the widespread (and presumably universal) tendency to avoid the use of formally (near-)identical and (near-)adjacent (non-

coordinate) grammatical elements or structures". That is, language users tend to avoid using two (or more) successive *to*-infinitives or *–ing*-forms, especially when the succession is immediate. The *horror aequi* principle, therefore, affects the complementation of the matrix predicate – sometimes with unpredicted and rather peculiar results, as can be seen in the following example from Fanego (2007, 177):

> (23) …it was not safe for me *to attempt doing* him any good. (Defoe, *Roxana*, 1724)

The example complies with Rohdenburg's (ibid.) "avoidance strategies" which suggest that in order to avoid an undesirable sequence of marked infinitives, the introduction of the second *to*-infinitive could be delayed or just replaced with another verbal or prepositional alternative altogether.

2.5.4 The Great Complement Shift

Another factor which has a bearing on the choice of complement is the phenomenon known as the Great Complement Shift (Rohdenburg 2006, 143). Defined as a "system of changes affecting the complementation system of English" (Rudanko 2006, 46), it is an ongoing development which is thought to have been in motion already since the Middle English period (Fanego 1996a, 33). The most significant change caused by this shift is the increased preference of the gerund (*–ing*-form) over the *to*-infinitive, which can be seen in the examples below (Rohdenburg 2006, 143-144):

> (24) She delighted to do it. $\rightarrow$ She delighted in doing it.
> (25) She was used/accustomed to do it. $\rightarrow$ She was used/accustomed to doing it.
> (26) She avoided/dreaded to go there. $\rightarrow$ She avoided/dreaded going there.

It has been noted by Vosberg (2009, 214-226), among others, that ever since the mid nineteenth century, American English has welcomed these changes more eagerly whereas British English is found to be more conservative and in many cases seems to be lagging behind its transatlantic cousin in this direction of development. This is not a very relevant fact to us as the present thesis does not

concern American English; however, there may be some evidence of the delay with which British

English accepts gerundial complements instead of infinitives in the data.

3 On corpora

In this chapter, I will introduce the corpora used in this thesis, as well as explain some key concepts that are essential in corpus linguistics.

3.1 The corpora used in this thesis

As mentioned in the introduction, the corpus section of this thesis is based on data from *The Corpus of Late Modern English Texts* (CLMET) and *The British National Corpus* (BNC).

3.1.1 *The Corpus of Late Modern English Texts*

The CLMET is compiled by Henrik De Smet and contains texts from the *Project Gutenberg* and the *Oxford Text Archive* (De Smet 2005, 70). The corpus is divided into three sections or "sub-periods", all of which cover a period of 70 years: 1710–1780, 1780–1850, and 1850–1920. These are referred to as CLMET 1, CLMET 2 and CLMET 3 respectively. The first part contains 2,096,405 words from 24 texts written by 15 authors, CLMET 2 consists of 3,739,657 words from 39 texts written by 29 authors, and CLMET 3 is made up of 3,982,264 words from 52 texts by 28 different authors. In total, therefore, the corpus contains 115 texts written by 72 different authors and the total number of words in the whole corpus is 9,818,326.

According to De Smet (ibid., 70-71), the collection of texts for the corpus has four principles. First of all, in order to increase the homogeneity of the texts within each sub-period and decrease the homogeneity between the different sub-periods, all the texts included in one sub-period must be written by authors who were born within the same restricted time-span. This principle also allows easier observation of historical trends. The sub-periods in question can be seen in the following figure (from De Smet 2005, 71):
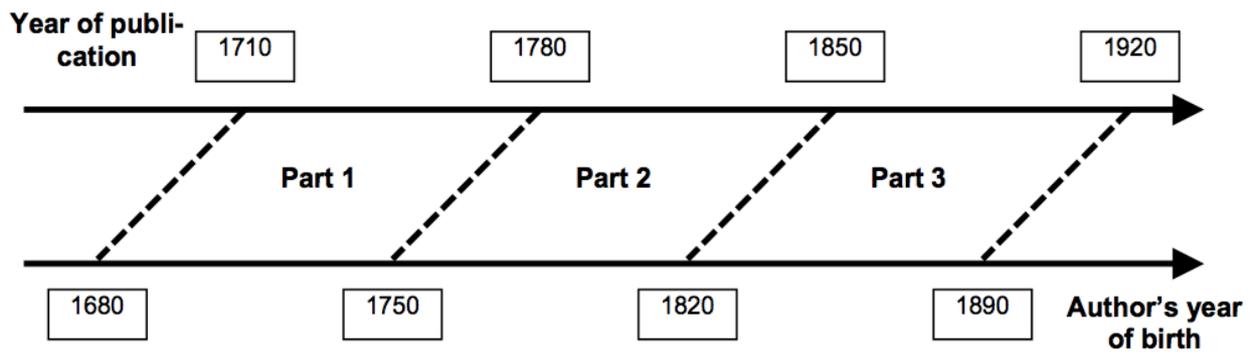
Figure 1: CLMET sub-sections.

Secondly, all the authors included in the corpus are native speakers of British English in order to avoid dialectal variation. Thirdly, the amount of text that can be included in the corpus by any one author is limited to 200,000 words per author, in order to "avoid thwarting of the data by the idiosyncrasies of individual authors" (ibid., 71). The fourth and last principle is that De Smet has paid special attention to include texts from different genres and from authors of different social backgrounds to ensure variation in the corpus. This has been done by favouring non-literary texts from lower registers and by lower-class authors, and including as many women writers as possible, to counterbalance the otherwise male-dominated and upper-class literary field of the 18th and 19th century.

3.1.2 *The British National Corpus*

The BNC consists of material created between the years 1960 and 1993, and therefore provides a modern-day viewpoint to the usage of *sorry* in contrast to the historical data of the CLMET. According to the BNC reference guide (2007), the corpus contains 96,986,707 words, of which 90 per cent is written material and the remaining 10 per cent spoken. However, this thesis will only focus on the Imaginative Prose section of the corpus which is made up of 16,496,408 words in total.

Three selection features affected the selection of material for the corpus: domain (subject

field), time (when published) and medium (book, periodical etc.) (ibid.). This was done in order to ensure that the corpus would include a diverse range of language style and usage, and that it would be "a microcosm of current British English in its entirety" (ibid.). The material has been gathered, for example, from books, newspapers, brochures, advertisements, letters and essays – in other words, all kinds of published and unpublished material.

3.2 Normalised frequencies

It is important to pay attention to the fact that the corpora used in this thesis are not equal in size and the results, therefore, are not directly comparable. This difference can be illustrated with an example (Biber et al. 1998, 263): if there are two texts that both include 20 occurrences of a specific word or construction, it would seem that the phenomenon in question is equally frequent in both texts. However, if we take into account that one of the texts is only 750 words long and the other consists of 1,200 words, it becomes obvious that the phenomenon is actually more frequent in the former text. It is important, therefore, to normalise the frequencies of the tokens so that the results can be compared between different corpora.

Normalised frequencies can be calculated by first dividing the number of occurrences that are being studied by the total number of words in the whole corpus. The result is then multiplied by a number that has been chosen for norming. In this study, that number will be 1,000,000, which should give us the right idea of the actual frequencies. The process is thus as follows:

$$\frac{20}{2,096,405} \times 1,000,000 = 9.54$$

The example above shows the calculation of 20 occurrences found in the first part of the CLMET, which consists of 2,096,405 words. The resulting NF (normalised frequency) is 9.54 words per million.

4 *Sorry* in dictionaries and grammars

Before we move on to corpus analysis, it is important to define the word we are dealing with in order to gain full understanding of its meanings and uses. In this chapter, therefore, I will first examine the definitions of *sorry* in some well-known dictionaries and then discuss its appearance in three carefully selected grammars of English.

4.1 *Sorry* in dictionaries

The dictionaries consulted for this thesis were the *Oxford English Dictionary*, the *Oxford Advanced Learner's Dictionary*, and the *Collins Cobuild Advanced Learner's English Dictionary.*

4.1.1 *Sorry* in *The Oxford English Dictionary*

There are three separate entries under *sorry* in *The Oxford English Dictionary* (hereafter referred to as the *OED*). The first one is a variant of the noun *sirrah*, meaning 'a familiar or (occasionally) contemptuous form of address to a man or a boy' (now regional and colloquial). As this thesis focuses on the adjective meaning of the word, this meaning will not be discussed any further. Similarly, the third entry – in which *sorry* is used as a verb with the meaning 'to grieve, to sorrow; to feel sorry *for*' (as in *I sorried for her*) – and its additional obsolete subsenses lie outside the scope of the present thesis and will therefore be ignored.

The second entry, however, is the one that is of interest to us. According to the *OED*, the roots of the adjective *sorry* as we know it go back to the Old English forms *sarg-* or *sarig* and the separately developed Early Middle English form *sorj*. It is related to or formed similarly with, for example, the West Frisian word *search* meaning 'sore, spotty, scabby' and the Old Saxon word

sērag 'sad'. The effect of these "relatives" can still be seen in Modern English – indeed, the word *sore* has the same Germanic linguistic ancestors as *sorry.*

The *OED* divides this meaning of *sorry* further into five subsenses, some of which are quite similar and therefore difficult to distinguish from each other. However, in order to avoid accidentally overlooking some aspect that might prove important or useful in the later stages of this thesis, I have decided not to omit anything and include all the relevant senses – despite their similarity – in the table which can be found below. Only the attributive senses (i.e. senses in which *sorry* is only used as a modifier in a noun phrase, e.g. *It will be a sorry day for the republic*) have been removed as they do not concern the present thesis. The complementation patterns in the right-hand column are my own addition based on the examples provided by the dictionary and are there in order to make it easier to compare the patterns.

OED entry	Example(s)	Pattern(s)
1. a. Distressed, sad; feeling grief or sorrow. Now *rare* except as merged with senses 2 and 3a.	**1722 D. DEFOE** I was in the greatest Confusion imaginable..; and began to be at odds with myself whether to be glad or *sorry*.	*sorry* + Ø
2. In predicative use, usually following a verb as *to be, to look, to seem,* etc.; grieved or vexed about a particular thing; regretful. Sometimes difficult to distinguish from sense 3a. **a.** With *about, at, for, †of,* etc.	**2001 R. TOPE** He hadn't seemed *sorry* about Sean's death. He'd been more upset at the disappearance of his fork. **1806 *Ann. Rev.*** We are *sorry* at observing references to Bryant. **1980 S. W. MARTINS** The estate was again *sorry* at the prospect of his leaving, but this time they thought someone was interested in taking on the farm. **1920 E. WHARTON** He was not *sorry* for the dénouement of his visit.	*sorry* + *about* + NP *sorry* + *at* + *ing*-form *sorry* + *at* + NP *sorry* + *for* + NP
b. With subordinate clause.	**1823 H. W. LONGFELLOW** I am rather *sorry* that the Exhibition falls so late in the year.	*sorry* + *that*-clause

c. With infinitive.	**1929 O. NASH** I'm *sorry* to report that my arm is wide awake and seems likely to remain so for some time.	*sorry* + *to*-infinitive
3. Feeling or expressing remorse; penitent; apologetic. Freq. with clause, infinitive, or adverbial phrase giving the reason for such remorse or apology. **a.** In predicative use following *to be, to look*, etc.	**1791 A. RADCLIFFE** I'm *sorry* I frightened you so last night. **1846 DICKENS** 'I'm *sorry* to give you so much trouble, Towlinson.' **1863 G. MEREDITH** Did you say in it you are *sorry* for your virulent offensive letter that I received? **1989 T. CLANCY** Okay, maybe I shouldn't'a snapped the one over the rail – lost my cool, and I shouldn't have – okay, I'm *sorry* about that. **2006 I. LAWRENCE** 'Well, I'm *sorry*,' she said, though she didn't *look sorry*, or *sound sorry*.	*sorry* + *that*-clause *sorry* + *to*-infinitive *sorry* + *for* + NP *sorry* + *about* + NP *sorry* + Ø
b. With ellipsis of verb: 'I am sorry'.	**1954 W. FAULKNER** He said, 'All right. *Sorry*. I didn't know you had a wife.'	*sorry* + Ø
c. Interrogatively, requesting the repetition of words that the speaker failed to hear or understand (or sometimes, hard to believe). **(a)** Used alone.	**1965 W. TREVOR** He said *sorry* meaning pardon.	*sorry* + Ø
(b) In ***I'm sorry***.	**2000 K. ATKINSON** 'I'm *sorry*?' I said, thinking I must have mis-heard.	*sorry* + Ø
5. a. In predicative use with *for* and complement, following *to be* or *to feel*: having sympathy or pity.	**1977 J. JOHNSTON** 'What teacher do you have?' 'Miss McCabe.' 'God, is she still going. I'm *sorry* for you.'	*sorry* + *for* + NP
b. ***sorry for oneself***: in a state of self-pity; depressed, miserable. Now chiefly in ***to feel sorry for oneself***.	**1973 R. M. BROWN** I'm not sittin' here on this Goddamn subway train feeling *sorry* for myself.	*sorry* + *for* + NP$_{REFL}$

Table 2: meanings given to the adjective sorry *by the OED.*

As can be seen from the table, sense number two can take various different complements, including prepositional phrases (*about, at, for*) and sentential complements (*that*-clauses, *to*-infinitives, *at + – ing*-form). In most of the other senses, however, *sorry* is used as a predicative adjective without a complement (*sorry + Ø*). Because this thesis deals with complementation, this construction does not play a significant role in later discussion and analysis, as it does not include any complements. Nevertheless, it may be mentioned occasionally due to its frequent occurrence.

4.1.2 *Sorry* in *Oxford Advanced Learner's Dictionary*

The Oxford Advanced Learner's Dictionary (*OALD*) divides the definition of *sorry* into four separate senses, the fourth of which is the attributive sense and has therefore been omitted. The first three senses are illustrated below, accompanied with examples (also from the *OALD*):

(1) '[not before noun] feeling sad and sympathetic'
I'm sorry that your husband lost his job.

(2) '[not before noun] feeling sad and ashamed about sth that has been done'
We're very sorry about the damage to your car.

(3) '[not before noun] feeling disappointed about sth and wishing you had done sth different or had not done sth'
She was sorry that she'd lost contact with Mary.

In addition to the three senses mentioned above, the *OALD* lists the idiomatic constructions *be/feel sorry for sb* ('to feel pity or sympathy for sb'), *feel sorry for yourself* ('to feel unhappy; to pity yourself'), *I'm sorry* ('1: used when you are apologizing for sth; 2: used for disagreeing with sb or politely saying 'no'; 3: used for introducing bad news') and *I'm sorry to say* ('used for saying that sth is disappointing') as their own entries. The interrogative sense (sense 3c in the *OED*) is also mentioned.

It seems that the first sense mentioned in the *OALD* corresponds quite well with sense 2 in the *OED*, and the second sense with sense 3a. The third sense, 'feeling disappointed about sth and wishing you had done sth different or had not done sth', however, is vaguer in meaning and could

actually be paired with either sense 2 or sense 3a, depending on the context.

When the complementation patterns given for the definitions in the two dictionaries are compared, it becomes even clearer that trying to make the definitions match is not simple. The *OALD* mentions only two patterns for the first sense: *~ (that...)* and *~ (to see, hear, etc.)*, which differs considerably from the definition of the corresponding sense 2 in the *OED*, where altogether six different patterns can be recognised. The same is true with the second sense, which only lists *~ (for/about sth)* and *that*-clause as possible complements, while the corresponding sense 3a in the *OED* mentions *to*-infinitives, too. The *OALD*'s third sense can take both *that*-clauses and *to*-infinitives as complements, but not prepositional phrases, which is the case with both sense 2 and 3a in the *OED*.

4.1.3 *Sorry* in *Collins Cobuild Advanced Learner's English Dictionary*

Cobuild divides the definition of *sorry* into ten senses. The definition of the first sense is

> You say 'Sorry' or 'I'm sorry' as a way of apologizing to someone for something that you have done which has upset them or caused them difficulties, or when you bump into them accidentally.

Again, the definition differs from the ones in the *OED* and *OALD*, but the sense seems to correspond quite well with sense 3a in the *OED*. The examples in *Cobuild* show that this sense can take at least *that*-clauses, *to*-infinitives, and prepositional phrases beginning with *for* as complements, even though the patterns themselves are not marked next to the definition.

The second sense is described with the words "If you are sorry about a situation, you feel regret, sadness, or disappointment about it" with three complementation patterns mentioned. The first is [v-link ADJ], in which *v-link* stands for *link verb*, such as *be*. In other words, this represents the *sorry* + Ø pattern. They also mention [ADJ *about* n] and [ADJ that/to-inf] as possible patterns – note that *Cobuild* uses "n" instead of "NP" to denote noun phrases. This sense could be quite easily compared to sense 2 in the *OED*.

The next noteworthy sense is defined as "If you feel sorry for someone who is unhappy or in an unpleasant situation, you feel sympathy and sadness for them" with pattern [v-link ADJ *for* n]. This sense, too, could be classified as representing sense 2 in the *OED*. It is followed by a separate sense involving the construction *sorry + for + *NP$_{REFL}$ (corresponding with sense 5b in the *OED*) with a rather lengthy definition: "You say that someone is feeling *sorry for themselves* when you disapprove of the fact that they keep thinking unhappily about their problems, rather than trying to be cheerful and positive". In *Cobuild*, this complementation pattern is marked with [v-link ADJ *for* pron-refl], and also labelled "disapproval".

In addition to the four senses mentioned above, *Cobuild* also lists the phrase *I'm sorry to say* as its own sense, with the definition "You use the expression *I'm sorry to say* to express regret together with disappointment or disapproval". This is marked with the label "phrase" and the complements mentioned next to it – [PHR with cl] and [PHR that] – refer to the complements of the whole phrase rather than the word *sorry* itself. The five remaining senses correspond with senses 1 and/or 3 in the *OED*, and no complementation patterns are mentioned with these senses.

It is clear that the definitions in *Cobuild* are, once again, quite different to the ones in the *OED*. However, both dictionaries list the same complements for *sorry*, with only one difference: *Cobuild*, similarly to the *OALD*, does not mention prepositional phrases beginning with *at* as possible complements. This might imply that this construction is somewhat rarer than the others which were mentioned in all three dictionaries.

4.1.4 Summary of senses

Since the senses and their definitions in the three dictionaries are so different, I thought it would be useful for the purposes of this study to first gather up all the different senses and then divide them again so that only a few simplified meaning groups remain. The table below presents a simplified and condensed summary of the different senses of *sorry* based on the dictionary findings above.

Simplified sense	Meaning	Patterns
1. Polite	Expressing sympathy and/or sadness for something that has happened (to someone else); expressing sympathy for **someone.**	*sorry + about* + NP *sorry + for* + NP *sorry + that*-clause *sorry + to*-infinitive
2. Sadness, Disappointment, or Regret	Expressing sadness, disappointment and/or regret about something that has happened (that affects the speaker themselves).	*sorry + about* + NP *sorry + at + ing-form* *sorry + at + NP* *sorry + that*-clause *sorry + to*-infinitive
3. Apology	Apologising for something that has happened.	*sorry + about* + NP *sorry + for* + NP *sorry + that*-clause *sorry + to*-infinitive *sorry + Ø*
4. Self-pity	Feeling unhappy or pitying yourself.	*sorry + for* + NP_{REFL}

Table 3: Summarised senses of sorry.

As can be seen from the table, the different definitions in the dictionaries have been simplified so that only four senses remain. First of all, sense 1 marks a 'polite' *sorry*; the speaker is not the cause of the regrettable event, but wishes to express their condolences anyway. Sense 2 is sometimes difficult to distinguish from sense 1, but it generally depicts disappointment rather than sympathy for something (e.g. *OED*'s example *I am rather sorry that the Exhibition falls so late in the year*). In sense 3, however, the speaker themselves has indeed done something that they must apologise for. Finally, sense 4 represents the *sorry* + NP$_{REFL}$ pattern (meaning 'self-pity'), which all the dictionaries classified as a separate subsense and not belonging under the construction *sorry + for +* NP.

Looking at the table, it seems peculiar that the *sorry + Ø* pattern (*I'm sorry*) only occurs in the third sense, i.e. when the speaker is apologising for something that he or she has caused. To me, it would seem perfectly fine to say *I'm sorry* in a sympathetic way, in a situation where I am not to blame for the matter that is being discussed. However, this use was not mentioned in any of the dictionaries – not in any easily distinguishable way at least. The corpus analysis section of this

thesis will hopefully shed some light on whether this use actually is somehow unacceptable or unheard of, or merely ignored because of its obviousness.

4.2 *Sorry* in grammars

For this section, I consulted three widely used reference grammars: *The Cambridge Grammar of the English Language* (2002) by Rodney Huddleston and Geoffrey K. Pullum, *A Comprehensive Grammar of the English Language* (1985) by Randolph Quirk et al., and *Longman Grammar of Spoken and Written English* (1999) by Douglas Biber et al.

Biber et al. (1999, 672) classify *sorry* as an adjective belonging to the group of "affective adjectives" (as opposed to "certainty adjectives"), as do words such as *afraid, astonished, disappointed, glad, (un)happy* and *sad*. They can take post-predicate *that*-clauses as complements and they describe "a personal attitude or feeling towards the proposition in the *that*-clause". Biber et al. also provide an example (ibid., 672):

(1) I'm *sorry* I hit you just now.

Quirk et al. (1985, 1222) also mention *that*-clauses, but define their use even further by stating that *sorry* can have either an indicative verb or putative *should* in the following *that*-clause. The indicative verb "refers to an established fact", whereas putative *should* follows expressions of emotion and often occurs with intensifying expressions. Compare these examples from Quirk et al. (ibid.):

(2) I am *sorry* (that) I have to leave so early.
(3) I am *sorry* (that) you *should have* been (so) inconvenienced.

In addition, Biber et al. (ibid., 716) note that "adjectives of personal affective stance" (a group which I interpreted as including the same adjectives as the "affective adjectives" mentioned above) can control post-predicate *to*-clauses. They go on to explain that in a situation where *to*-clause is preceded by an adjective of personal affective stance, such as *sorry*, the subject of the subordinate

clause is understood to have the same subject as the main clause (ibid., 719):

(4) I'm sorry *to hear about you*.

Sorry with prepositional phrases as complements is mentioned by both Huddleston & Pullum and Biber et al. The former list it generally as an adjective that can take complements with the form preposition + NP (Huddleston and Pullum 2002, 964), but also mention one specific preposition, *for*, that can occur with *sorry* (ibid., 544). The latter mention the construction *sorry + about + –ing*-form when discussing adjectival predicates that allow *–ing*-clauses (Biber et al. 1999, 749).

5 Corpus analysis

It is now time to move on to the analysis of the corpus data. In the following sections, I will first examine and discuss the data collected from the three parts of the CLMET, and then proceed to studying the sample taken from the BNC.

5.1 *Sorry* in CLMET 1: 1710-1780

The search for *sorry* in the first part of the CLMET yielded 152 tokens. 12 tokens (7.9%) were discarded due to the fact that *sorry* was used as a modifier in a noun phrase:

(1) Never did *sorry* traveller make such a pother and racket about his remarks as I did about mine, upon the occasion. (Sterne, *Life and Opinions of Tristram Shandy*, 1759-1767)

This left 140 usable tokens, the distribution of which is presented in the table below.

Pattern	No. of tokens	% of data	NF/million
sorry + *that*-clause	39	27.9	18.60
sorry + *to*-infinitive	50	35.7	23.85
sorry + *for* + NP	29	20.7	13.83
sorry + *for* + *wh*-clause	5	3.6	2.39
sorry + Ø	17	12.1	8.11
Total	**140**	**100**	**66.78**

Table 4: Complementation patterns found in CLMET 1.

As can be seen from the table, there were five different complement patterns in this part of the corpus. They will be studied more closely in the following sections, starting from the non-sentential *for* + NP, then moving on to the sentential alternatives, and discussing the zero complement last.

5.1.1 *For* + NP

The *for* + NP pattern could be found in 27 tokens (19.3%). In 20 of them, the subject was *I*, and all the other subjects had the semantic features +human and +animate, too: *you, we, they, one, the man*

and *the master of the hotel* were found in subject position in addition to *I*. Thus, in all the tokens, the subject has the thematic role of experiencer. The only example with the generic *one* as subject can be seen below.

> (2) … if they were out of the question, one could not be *sorry* for such a mortification to the pride of old Somerset. (Walpole, *Letters [Vol. 1]*, 1735-1748)

As for the NP inside the prepositional phrase, in 23 of the 27 tokens (85%) it was classified as both –animate and –human. Short pronouns such as *that* and *it* were common in this group, but there were also examples of longer NPs depicting something negative and unpleasant, such as *our fault, your misfortune* and *all the trouble.* Below are illustrations of both types.

> (3) I told the Christian reader – I say Christian – hoping he is one – and if he is not, I am *sorry* for it – and only beg he will consider the matter with himself, and not lay the blame entirely upon this book… (Sterne, *The Life and Opinions of Tristram Shandy, Gentleman*, 1759–1767)
>
> (4) "I am very *sorry*, sir, for your unhappiness, whatever is the occasion of it." (Fielding, *Tom Jones*, 1749)

The remaining four NPs were mostly proper nouns, but there was one third person pronoun, too. Again, examples of both types can be found below.

> (5) I cannot be *sorry* for Lord Granville, for he certainly sacrificed everything to please the King; … (Walpole, *Letters [Vol. 1]*, 1735-1748)
>
> (6) 'And is not my sister an angel, now, pappa,' cried the eldest, 'and why then are you *sorry* for her?' (Goldsmith, *The Vicar of Wakefield*, 1766)

All tokens with a +human NP in the prepositional phrase were interpreted to include *sorry* in the first sense ('sympathy'). However, in the –human group, the variation was greater: while most tokens represented sense 1 of *sorry*, there were occurrences of senses 2 and 3, too. Examples of all three senses together with a –human and –animate NP can be seen below.

> (7) We are heartily *sorry*, Captain, for your Misfortune. (Gay, *The Beggar's Opera*, 1728)
>
> (8) I was soon *sorry* for the warmth with which I had spoken this… (Goldsmith, *The Vicar of Wakefield*, 1766)
>
> (9) 'Indeed, Miss Jenny, we are *sorry* for our fault, and will follow your advice; which we now see is owing to your goodness.' (Fielding, *The Governess*, 1749)

There were two examples of extraction in the data, both of which represented relativisation:

> (10) …deprived of all intercourse with rational creatures; a circumstance for which I was not *sorry*, as I had the more time to project schemes of revenge… (Smollett, *The Adventures of Peregrine Pickle*, 1751)
>
> (11) Sophia has been ill with a cold; her head is to be dressed French, and her body English, for which I am *sorry*; her figure is so fine in a robe… (Walpole, *Letters [Vol. 1]*, 1735-1748)

Complexity factors were also present in five tokens which featured insertions between the adjective and preposition. In most cases, the intervening material consisted of just one word (*Captain, sir, too*), but there was one token in which the insertion was made up of three words.

> (12) So I bid La Fleur tell the master of the hotel, that I was *sorry* on my side for the occasion I had given him… (Sterne, *A Sentimental Journey Through France and Italy*, 1768)

Here we can also witness the effect of the *horror aequi* phenomenon for the first time: *that* has been omitted from the second *that*-clause (*the occasion [that] I had given him*).

5.1.2 *For* + *wh*-clause

The preposition *for* was found to occur together with a *wh*-clause, too. There were five tokens with this pattern and in all of them, the *wh*-clause was introduced with the word *what*. Again, the subjects in the tokens were +animate and +human and their semantic role was experiencer. In all five tokens with the *for* + *wh*-clause complement, *sorry* appeared in sense 3.

All the verbs in the complement clauses had past tense, and, in fact, the data consisted of only three different verbs: *do*, *say* and *happen.* Examples of all three uses are presented below.

> (13) Come, come, all's well, and I am *sorry* for what I have done. (Fielding, *Tom Jones*, 1749)
>
> (14) "What! Was she *sorry* for what she had said?" (Walpole, *The Castle of Otranto*, 1764)
>
> (15) His lordship, meeting us at our friend's house, declared he was *sorry* for what had happened; and that he had no intention to give umbrage. (Smollett, *The Expedition of Humphry Clinker*, 1771)

In addition to example (15), there was one other token in the data in which the verb *happen* was used in the subordinate clause. In all the other tokens, the subjects of the main and subordinate

clauses were the same. None of the tokens in this group deviated from the basic subject – verb –

object word order except example (10) above – due to the fact that it is an interrogative sentence.

5.1.3 *To*-infinitive

Sorry + to-infinitive was the most common pattern in the CLMET 1 data with 50 tokens (35.7%).

Again, the majority of the tokens had *I* as subject and the rest were of the +human type, too.

Interestingly, as many as 14 tokens included the construction *should be sorry + to*-infinitive:

> (16) I'm sure I should be *sorry* to affront any gentleman who has been so polite, and said so
> many civil things to me. (Goldsmith, *She Stoops to Conquer*, 1773)

I interpreted all tokens with this construction to represent sense 2 of *sorry*, as the possibility of

having future regrets or being disappointed is strongly implied. Thus, example (16) above could be

paraphrased as follows:

> (16') I'm sure I would regret affronting any gentleman who has been so polite…

> Insertions between *sorry* and the *to*-infinitive were found in four tokens. The intervening

material was a reporting clause in two tokens (*cries Amelia*, *I said*), an addressing term in one (*my*

dear lady), and a combination of both in one token, which is presented below.

> (17) 'I am *sorry*, my little dears,' said Mrs. Teachum, 'to give your tender bosoms the
> uneasiness I fear the contents of this letter will do… (Fielding, *The Governess*, 1749)

It seems that despite the cognitive complexity created by the inserted material, the author has

chosen (intentionally or unintentionally) not to favour the more explicit or 'bulkier' grammatical

construction in the following *that*-clause, but has in fact omitted the *that*, contrary to what the

complexity principle suggests.

> Extraction was found in three tokens, all of which represented the relative extraction type.

One of them is of particular interest:

> (18) …not only for his sake, but for the sake of the best woman in the world, whom we
> should be *sorry* to consider as yoked to a man of no worth nor honour. (Fielding,
> *Amelia*, 1751)

This example could be paraphrased as follows without greatly affecting its meaning:

> (18') …not only for his sake, but for the sake of the best woman in the world, whom we should be *sorry* to consider to be yoked to a man of no worth nor honour.

Here we can see that the *to*-infinitive construction in example (18) could be interpreted as having been affected by the *horror aequi* principle, and has therefore been replaced with a different construction altogether.

As for the thematic roles involved in the *to*-infinitive construction, the role of the main clause subject was experiencer without exception. However, there was more variation in the roles of the lower clause subjects – which, as was explained in section 2.4.1, can be marked with the null pronoun PRO. Altogether three different roles were found: experiencer, agent and theme.

The role of experiencer was the most common choice for a lower clause subject, as 27 of the 50 tokens represented this type. All the tokens which had the experiencer role in the lower clause were found to represent the second sense of *sorry* ('sadness, disappointment or regret'). Verbs *see*, *hear* and *find* (in the meaning of 'become aware' or 'realise') were particularly common in this group. Below are examples of all three verbs.

> (19) Our family does not exceed thirty-six; and I should be *sorry* to see the number augmented, as our accommodation won't admit of much increase. (Smollett, *The Expedition of Humphry Clinker*, 1771)
> (20) I am very *sorry* to hear that Lady Hervey is ill. (Chesterfield, *Letters to his Son*, 1746–1771)
> (21) "I should be *sorry*," cries Amelia, "to find I had conceived an affection for a bad woman… (Fielding, *Amelia*, 1751)

The number of lower clause subjects which took the role of agent was slightly smaller, 22. In this group, *sorry* appeared in either sense 2 or sense 3. The most frequent lower clause verb was *say*, which sometimes made interpreting the sense of *sorry* rather difficult, as it was hard to judge whether saying something simply made the speaker sad (sense 2) or whether the speaker was actually apologising for saying something (sense 3). Most tokens with *say* were interpreted to represent the former sense. However, there were exceptions such as the example below, which I

interpreted to denote actual apology:

> (22) She introduced all with– "I am *sorry* to say it; and it is friendship which bids me speak; and it is for their good it should be told you." (Fielding, *Amelia*, 1751)

Tokens with other verbs were usually somewhat easier to classify. Below is an example in which the lower clause subject was found to have the role of agent and in which *sorry* was interpreted to represent sense 3.

> (23) Dear George, I am *sorry* to renew our correspondence upon so melancholy a circumstance… (Walpole, *Letters [Vol. 1]*, 1735-1748)

Finally, there was one token in which the lower clause subject was found to take the thematic role of theme. In this token, the *to*-infinitive is in the passive form:

> (24) She felt no concern for the death of young Conrad, except commiseration; and she was not *sorry* to be delivered from a marriage which had promised her little felicity… (Walpole, *The Castle of Otranto,* 1764)

In this case, *sorry* was interpreted to denote sense 2.

5.1.4 *That*-clause

Sorry + *that*-clause was the second most common pattern in this data with 39 tokens (27.9%). A slight majority of them, that is 22 tokens, were of the implicit type in which *that* has been omitted. Once again, the most common subject by far was *I* and no –human or –animate subjects were found. However, there was one token in which the subject was the slightly more unusual *humanity*, displayed below.

> (25) "Shall humanity," says Mason, "be thankful or *sorry* that it is so?" (Walpole, *Letters [Vol. 1]*, 1735-1748)

On the other hand, it was found that the subject of the lower *that*-clause could also be –human and –animate. Of all 39 tokens, 30 tokens had a different subject in the main clause and the lower clause, and among these 30 tokens, five of the lower clause subjects were considered not to be human or animate. All of them belonged to the explicit *that* group. An example is displayed below.

(26) I am only *sorry* that the impertinence of any child of mine should have occasioned this exertion of your spirit… (Smollett, *The Expedition of Humphry Clinker*, 1771)

In this token, the subject expression of the *that*-clause is also rather lengthy and complex, which may be the reason why explicit *that* was chosen.

The overall number of tokens which had a different subject in the main and lower clauses (whether +animate or –animate) was proportionally larger in the explicit group, as 15 out of 17 tokens (88%) were of this type. In the implicit group, the number was 15 out of 22 (68%). Changing the subject between the clauses could therefore be seen as creating a more complex environment, which is why the explicit *that* is needed.

Based on the CLMET 1 data, insertions, too, seem to play quite a significant part in whether *sorry* selects an explicit or implicit *that*-clause as complement. In the explicit group, insertions between *sorry* and the complement were found in 7 tokens out of 17 (41%); in the implicit group, however, the number of insertions was considerably smaller, as only two tokens out of 22 (9%) were found to contain an insertion. In one of them, the insertion was just one short word (*Sir*) which does not create a particularly complex construction, but in the other one, the intervening material could – according to the complexity principle – be considered lengthy enough to require an explicit *that*:

(27) "I am very *sorry*, sir," cries Booth very gravely, and turning as pale as death, "you should entertain a thought of this kind; … (Fielding, *Amelia*, 1751)

In fact, the data contained another token from the same author which was quite similar to example (27) above (if not actually slightly *less* complex), but in which the author had indeed decided to use an explicit *that*:

(28) "I am *sorry*, young gentleman," answered the stranger, "that you have any reason to be so unhappy at your years." (Fielding, *Tom Jones*, 1749)

Example (27), therefore, could be seen as somewhat of an anomaly and conflicting with the complexity principle. Interestingly, example (17) discussed earlier in section 5.1.3 is also from the

same author, which is why I am tempted to label this tendency to prefer these more complex

constructions over bulkier alternatives complying with the complexity principle as an idiosyncracy

of this particular author, at least until more evidence of similar behaviour is found.

In addition, there was one token in which the explicit use of *that* may have been triggered by

an unusual word order.

(29) *Sorry* I am, that the task I have undertaken, lays me under the necessity of divulging
this degeneracy… (Smollett, *The Adventures of Peregrine Pickle*, 1751)

There was also one token in the implicit *that* group in which the removal of *that* could be a result of

the *horror aequi* effect. Consider:

(30) Call upon them often, though you do not stay with them so long as formerly; tell them
that you are *sorry* you are obliged to go away, but that you have such and such
engagements… (Chesterfield, *Letters to his Son*, 1746-1771)

In this example, the consecutive relative clauses introduced by *that* and a desire to avoid excess

repetition might be the reason why *that* has been omitted in the complement clause.

The *that*-clause complement was found to occur with senses 1, 2 and 3 of *sorry*. In the

explicit group, senses 1 and 2 both appeared in eight tokens out of 17 (47% + 47%), while sense 3

was only found in one (6%). However, in the explicit group, sense 3 alone accounted for eight

tokens out of 22 (36%), but senses 1 and 2 only had seven tokens each (32% + 32%).

5.1.5 Ø complement

There were 17 tokens (12.1%) in the CLMET 1 data which did not contain a complement at all. One

of them is presented below:

(31) …but being informed that we were almost killed by the fright, they were vastly *sorry*;
but hearing that we had a very good night, they were extremely glad again.
(Goldsmith, *The Vicar of Wakefield*, 1766)

While the majority of the tokens in this group are simple variations of the phrase *I am sorry* in

different environments, there are some recurring constructions which require closer attention.

In two tokens, *sorry* was followed by *when*:

> (32) In fine, lieutenant Lismahago is a curiosity which I have not yet sufficiently perused;
>   and therefore I shall be *sorry* when we lose his company… (Smollett, *The Expedition
>   of Humphry Clinker*, 1771)
> (33) … we acquire such an attention to this end, that we are very uneasy under any
>   appointments, and are *sorry* when we either miss our game, or fall into any error in
>   our reasoning. (Hume, *A Treatise of Human Nature*, 1738)

In addition, there were five tokens with the construction *I should be sorry if…*, two of which are

presented below.

> (34) I should be *sorry* if what is here recommended should be at all understood to
>   countenance a careless or indetermined manner of painting. (Reynolds, *Seven
>   Discourses on Art*, 1769-1776)
> (35) I'm sure I should be *sorry* if people said anything amiss, since I have no fortune but my
>   character. (Goldsmith, *She Stoops to Conquer*, 1773)

At first glance, the elements following *sorry* in these four examples would appear to be adjuncts

rather than complements. However, we can apply certain tests to make sure this is the case.

According to Huddleston and Pullum (2002, 221) and Somers (1984, 509), complements can be

distinguished from adjuncts by trying to eliminate them and judging whether the utterance still

makes sense. If the element can be omitted, it must be an adjunct or an optional complement, as was

discussed earlier in sections 2.1 and 2.2 of this thesis. Examples (32) and (33) would look as

follows if the *when*-clauses were removed:

> (32') In fine, lieutenant Lismahago is a curiosity which I have not yet sufficiently perused;
>   and therefore I shall be *sorry* ~~when we lose his company…~~
> (33') … we acquire such an attention to this end, that we are very uneasy under any
>   appointments, and are *sorry* ~~when we either miss our game, or fall into any error in
>   our reasoning.~~

It is clear that the examples seem to lose something that is very essential to the meaning of the

sentence when the *wh*-clauses are eliminated. Indeed, the same is true with the *if*-clauses:

> (34') I should be *sorry* ~~if what is here recommended should be at all understood to
>   countenance a careless or indetermined manner of painting.~~
> (35') I'm sure I should be *sorry* ~~if people said anything amiss, since I have no fortune but
>   my character.~~

However, Huddleston and Pullum also state that adjuncts can be moved around in a sentence more

freely than complements (ibid., 225). The *wh*-clauses appear to respond quite well to being moved into another position within the sentence (extracted element underlined):

> (32'') In fine, lieutenant Lismahago is a curiosity which I have not yet sufficiently perused; and therefore, <u>when we lose his company,</u> I shall be *sorry* …
>
> (33'') … we acquire such an attention to this end, that we are very uneasy under any appointments, and <u>when we either miss our game, or fall into any error in our reasoning</u>, [we] are *sorry*.

This seems to be the case with the *if*-examples, too:

> (34'') <u>If what is here recommended should be at all understood to countenance a careless or indetermined manner of painting</u>, I should be *sorry*.
>
> (35'') <u>If people said anything amiss</u>, I'm sure I should be *sorry*, since I have no fortune but my character.

In addition, the word *if* could be replaced with a construction such as *in the event that* without changing the meaning of the utterance or making it ungrammatical. This evidence suggests that the *when*- and *if*-clauses in these examples indeed behave more like adjuncts than complements. However, they do seem to be closely connected to the adjective in the main clause, because even though the attempt to eliminate them does not result in ungrammaticality as such, the sentence loses an element which is vital to the understanding of its message. Therefore, they will be referred to as "adjunct-like" elements from now on.

As for the different senses of *sorry* found in this complement group, sense 2 was the most common appearing in 10 tokens – including all the tokens with the expression *I should be sorry if…*, as could be expected based on the findings presented in section 5.1.3. However, there were also seven tokens in which an apology was clearly meant and which therefore represented sense 3. Below is an example of the latter.

> (36) "If I have offended you, madam," said Amelia, "I am very *sorry*, and ask your pardon; … (Fielding, *Amelia*, 1751)

5.1.6 Review of CLMET 1

Sorry was found to occur with altogether five different complement types in this data from 1710–1780. *To*-infinitive was the most common choice with a 35.7% share of all the 140 relevant tokens. A particularly frequent phenomenon amongst the *to*-infinitive tokens was the *should be sorry + to*-infinitive construction, which appeared 14 times. It was also found that the thematic role of the lower clause subject could be either experiencer, agent or theme. If it was experiencer, *sorry* took sense 2, but if the role was agent, both senses 2 and 3 were possible. In the only token in which the role of the lower clause subject was theme, the verb was passivised and *sorry* took sense 2.

That-clause was the second most common complement with 39 tokens, or 27.9%. The majority were of the implicit type in which *that* had been omitted. The relation between the subjects of the clauses seemed to have an effect on the complement choice, as 88% of the explicit tokens had different subjects in the main and lower clause. In addition, complexity factors in the form of insertions and extractions were much more frequent in the explicit group. *Sorry* took either sense 1, 2 or 3 with a *that*-clause complement.

Complements with the preposition *for*, *for* + NP and *for* + *wh*-clause, were clearly in the minority. With the former, it was found that if the NP inside the prepositional phrase had the semantic features +human and +animate, *sorry* took sense 1. The most remarkable finding about the *for* + *wh*-clause complement was that only three different verbs, *do*, *say* and *happen,* appeared in the *wh*-clause.

In addition, there were 17 tokens (12.1%) in which *sorry* did not take any complement. However, it was found to occur frequently with *if*- and *when*-clause adjuncts.

5.2 *Sorry* in CLMET 2: 1780-1850

The second part of the CLMET contains texts published between the years 1780 and 1850. The search string *sorry* returned 424 tokens in this part of the corpus, of which 50% – that is, 212

tokens – were randomly selected for analysis. In 11 tokens out of 212 (5.2%), *sorry* was either

– a modifier in a noun phrase:

(1) … I mounted a *sorry* mule without bridle or stirrups … (Borrow, *The Bible in Spain*, 1843)

– or not used as an adjective:

(2) "Then make that *sorry* of use to you, Cecilia, and fix it steadily in your thoughts …" (Edgeworth, *The Parent's Assistant*, 1796)

All of these were discarded and will not be analysed further because they lie outside the scope of

the present thesis. This left me 201 usable tokens. The table below illustrates the distribution of

complements in the remaining data.

Pattern	No. of tokens	% of data	NF/million
sorry + *that*-clause	53	26.4	28.34
sorry + *to*-infinitive	71	35.3	37.97
sorry + *at* + *–ing*-form	1	0.5	0.53
sorry + *for* + *–ing*-form	1	0.5	0.53
sorry + *for* + *wh*-clause	1	0.5	0.53
sorry + *at* + NP	2	1	1.07
sorry + *for* + NP	40	19.9	21.39
sorry + Ø	32	15.9	17.11
Total	**201**	**100**	**107.47**

Table 5: Complementation patterns found in CLMET 2.

Altogether eight different complement patterns were found in the data, including the Ø

complement. Again, these will be discussed more thoroughly below.

5.2.1 *For* + NP

The prepositional pattern *sorry* + *for* + NP was the most common non-sentential complement with

40 occurrences. While the subject was mostly *I* or some other personal pronoun, the NP inside the

prepositional phrase was –animate and –human in 26 cases (most of which were either *it, that* or

this), +animate and +human in 15 cases, and +animate, –human in one token. Below are examples

of all three types.

> (3) "I'm *sorry* for it, Miss Catherine," was my response; and I proceeded assiduously with
> my occupation. (Brontë, *Wuthering Heights*, 1847)
> (4) I was *sorry* for the murderer, whoever he might be… (Gaskell, *Mary Barton*, 1848)
> (5) "How delighted he will be when he hears the good news!" said she, to herself; "but I
> know he will be a little *sorry* too for my poor lamb." (Edgeworth, *The Parent's
> Assistant*, 1796)

Again, in all the tokens with +animate NP, *sorry* took sense 1. If the NP was –animate, sense 1 was

still the most common one, but there were occurrences of senses 2 and 3, too.

Extractions were found in two tokens in this sample, both of which were relativisations.

Interestingly, both of them were also from the same source. The tokens in question are presented

below.

> (6) …but neither he nor my mother would hear of such a thing, which I was very *sorry* for,
> as it would have been so convenient to me… (Galt, *The Ayrshire Legatees*, 1821)
> (7) … in the proprieties of London life, say that it would be very vulgar in me to go to look
> at her, which I am *sorry* for, as I wish above all things to see a personage so
> illustrious by birth… (Galt, *The Ayrshire Legatees*, 1821)

Insertions between *sorry* and the preposition did not play a remarkable part in this sample, as only

four tokens had them, and they were all short words such as *enough* and *too*.

> (8) I'm *sorry* enough for the girl, for bad's come over her, one way or another, but I'm
> sorrier for my wife. (Gaskell, *Mary Barton*, 1848)

The clause containing the *sorry + for* + NP construction itself, however, was an insertion in three

tokens. In one of them, the copular verb had been omitted:

> (9) Mary, though pretending not to hear, was somewhat disconcerted; and Elizabeth, *sorry*
> for her, and sorry for her father's speech, was afraid her anxiety had done no good.
> (Austen, *Pride and Prejudice*, 1813)

5.2.2 *At* + NP

There were two complement types with the preposition *at*: *at* + NP and *at* + *–ing*-form. The former

occurred two times in the data. I interpreted both tokens with *at* + NP construction to represent

sense 2, and the subject has the thematic role of experiencer in both cases. In addition, the semantic

features of the NPs were –human and –animate in both tokens. As the number of the tokens is so

small, they are both presented here.

> (10) … because she thought so well of him, and desired still to think so well, that she was
> *sorry* at any faults which rendered him less worthy of her good opinion. (Inchbald,
> *Nature and Art*, 1796)
>
> (11) After a slight remark he passed on; and I was not *sorry* at his disturbance, as it enabled
> me to pass through Eton before people were generally up. (De Quincey, *Confessions
> of an English Opium Eater*, 1822)

The only dictionary in which the *at* + NP and *at* + *–ing*-form patterns were mentioned as possible

complements of *sorry* was the *OED*, which implies that they are nowadays rather uncommon.

Indeed, the preposition *at* does seem somewhat redundant, as it could easily be replaced with *for* –

or perhaps even *about* – in both examples. The third part of CLMET and the BNC sample will

hopefully shed more light on whether *at* is still a valid choice or whether it has been completely

replaced with other prepositions.

5.2.3 *For + –ing*-form

The *for + –ing*-form pattern occurred only once in this sample.

> (12) And I will find will; and then, Margaret, I think you'll be *sorry* for being so stubborn
> about Jem. (Gaskell, *Mary Barton*, 1848)

The reason why this particular pattern has been chosen could lie in the principle known as The

Great Complement Shift introduced in section 2.5.4. Consider the same token paraphrased with a

to-infinitive:

> (12') And I will find will; and then, Margaret, I think you'll be *sorry* to have been so
> stubborn about Jem.

The fact that the gerundial construction appeared only once in CLMET 2 data and 0 times in

CLMET 1, while the number of tokens containing the *to*-infinitive was 71 and 50 respectively,

seems to support this assumption. The gerund has emerged, but it is nowhere near as common as the

infinitival alternative. It is nevertheless very difficult to make any conclusions based on just one token, but in sections 5.3 and 5.4 we shall find out whether the use of the –ing-form has increased during the second half of the 19th century and the 20th century, as The Great Complement Shift theory would suggest.

5.2.4 *For + wh*-clause

The *for + wh*-clause pattern also occurred only once.

> (13) I daresay, however, when your passion is over, and when you recollect yourself, you are very *sorry* for what you have done and said; are not you? (Edgeworth, *The Parent's Assistant*, 1796)

In the first part of the CLMET corpus, only a few verbs seemed to occur in the *wh*-clause (*do, say* and *happen*). Here, the lower clause verb is again *do* and the subject is the same in both clauses, which means that *sorry* in this token quite obviously represents sense 3. In the main clause, the thematic role of the subject *you* is experiencer, while in the lower clause the same subject takes the role of agent.

5.2.5 *At + –ing*-form

At + –ing-form also belongs to the group of complements which occurred only once in the CLMET 2 data.

> (14) I was not *sorry* at being thus honoured to suffer in the cause of righteousness… (Hogg, *The Private Memoirs and Confessions of a Justified Sinner*, 1824)

Again, this token could be paraphrased as follows using a *to*-infinitive:

> (14') I was not *sorry* to be thus honoured to suffer in the cause of righteousness…

However, before declaring this token to be another example of The Great Complement Shift, the proximity of the *to*-infinitive in the following clause (*to suffer*) makes me suspect that the *horror aequi* principle may have also had an effect on the choice of the gerundial complement in this case.

5.2.6 *To*-infinitive

The *to*-infinitive pattern was the most frequent complement in CLMET 2, too, with 71 tokens

(35.3%). Again, as many as eight tokens featured the construction *should be sorry + to*-infinitive:

> (15) "No, Lizzy, that is what I do NOT choose. I should be *sorry*, you know, to think ill of a
> young man who has lived so long in Derbyshire." (Austen, *Pride and Prejudice*,
> 1813)

Interestingly, three of the tokens containing the *should* construction were from novels written by

Anne Brontë (*Agnes Grey* and *The Tenant of Wildfell Hall*), and two – including example (15)

above – from Jane Austen (*Pride and Prejudice* and *Sense and Sensibility*). The remaining three

tokens were by different authors. It seems, therefore, that Brontë and Austen had a particular

tendency to use this construction.

Insertions between *sorry* and the *to*-infinitive appeared in six tokens. This pattern seems to

allow particularly long insertions as opposed to, for example, the *for* + NP pattern discussed above

which only had short words such as *enough* and *too* as insertions. Below are two examples of these

long insertions found in this sample.

> (16) I so much enjoyed my rambles among the rocks and mountains of St. Helena, that I felt
> almost *sorry* on the morning of the 14th to descend to the town. (Darwin, *The Voyage
> of the Beagle*, 1839)
> (17) It grew rather late before my kind hosts had finished their narrations, and I was not
> *sorry*, after all the exercise I had taken, to return to my cell, where everything invited
> to repose. (Beckford, *Dreams, Waking Thoughts, and Incidents*, 1783)

Extraction was found in one token, which is presented below.

> (18) …a very good sort of man, indeed; but from whom I was not at all *sorry* to be
> delivered. (Beckford, *Dreams, Waking Thoughts, and Incidents*, 1783)

In the CLMET 2 sample, the lower clause subject was more likely to take the thematic role

of agent rather than experiencer, as 38 tokens out of 71 were of this type. Again, the tokens in

which the subject of the lower clause had the role of agent were found to represent either sense 2 or

3 of *sorry*. Here are examples of both senses:

(19) As I have been most delighted with the country parts of Norway, I was *sorry* to leave Christiania without going farther to the north… (Wollstonecraft, *Letters on Sweden, Norway and Denmark*, 1796)

(20) Gentlemen, I am *sorry* to interrupt your amusements… (Edgeworth, *The Parent's Assistant*, 1796)

The subject of the lower clause was found to have the role of experiencer in 28 cases, all of which also featured *sorry* in the second sense. Again, the verbs *see*, *hear* and *find* were very common in this group. Below is an example of the rarer alternatives.

(21) But your lordship, I am *sorry* to observe to-day, is troubled with the gout… (Byron, *Letters*, 1810–1813)

Additionally, there were four tokens in which the lower clause subject took the role of theme. Again, the *to*-infinitive in these tokens was in the passive form and all of them were classified as tokens representing sense 2.

(22) I was to have dined at Deane to-day, but the weather is so cold that I am not *sorry* to be kept at home by the appearance of snow. (Austen, *Letters to her Sister Cassandra and Others*, 1796-1817)

Finally, one token was found to feature the lower clause subject in recipient role. Consider the following example:

(23) I should be very *sorry* to receive a charming little lady-like note from any of them… (Brontë, *Agnes Grey*, 1847)

This token was also interpreted to represent sense 2 of *sorry*.

5.2.7 *That*-clause

Sorry + *that*-clause was the second most frequent pattern in the CLMET 2 data with 53 tokens (26.4%). In 21 tokens, the *that*-clause was of the explicit variety, while the majority – that is 32 tokens – featured the implicit form. The number of tokens which had different subjects in the main clause and *that*-clause was 14 in both groups; however, if we translate them into percentages, it becomes obvious that changing the subject between clauses was more frequent in the explicit group (67%) than the implicit one (44%). In both groups, the most common subject in the main clause was

again *I* or some other personal pronoun, but the subject of the *that*-clause could have the semantic features –human and –animate, too. In the explicit group, five out of 14 tokens were of this type, but in the implicit group, only one token had these features. Below are examples of both explicit and implicit types.

> (24) "I have always been swayed by your counsel," said I, "and for your sake, principally, I am *sorry* that all our measures have proved abortive. (Hogg, *The Private Memoirs and Confessions of a Justified Sinner*, 1824)
> (25) I am so *sorry* my hand is unintelligible; but I can neither deny your accusation, nor remove the cause of it. (Byron, *Letters*, 1810-1813)

As for the different senses of *sorry*, sense 2 was the most common sense in the explicit group with 47% of all the explicit tokens representing this type. Sense 1 accounted for 33% and sense 3 had a 20% share. In the implicit group, however, the proportions were quite different: in the majority of the tokens (47%), *sorry* took sense 3. Senses 1 and 2 appeared in 16% and 38% of the tokens, respectively.

As was the case in CLMET 1, insertions seem to have an effect on whether the explicit or implicit alternative is chosen. In the explicit group, insertions were found in three tokens, but there were none in the implicit group. Below are all the explicit tokens containing insertions.

> (26) I am *sorry*, if he be, that I have published his disgrace, for I would not hurt YOU. (Edgeworth, *The Parent's Assistant*, 1796)
> (27) Dolly was more *sorry* than she could tell, that he should have taken so much trouble. (Dickens, *Barnaby Rudge*, 1841)
> (28) I am heartily glad that you have escaped any share in the impurities of Deane, and not *sorry*, as it turns out, that our stay here has been lengthened. (Austen, *Letters to her Sister Cassandra and Others*, 1796-1817)

In addition, there was one token in which moving *sorry* to the front of the main clause (combined with the very heavy subject expression in the lower clause) could play a part in why the explicit complement was chosen.

> (29) *Sorry* am I that the shackles of modern decorum restrain me from penning that famous rebuke;… (Hogg, *The Private Memoirs and Confessions of a Justified Sinner*, 1824)

In the implicit group, on the other hand, there were at least five tokens which seem to have been affected by the *horror aequi* principle, that is, the close proximity of another *that*-clause. An example can be seen below.

> (30) …round his neck, as he stood leaning upon the sill of the window – 'or should I have said that you are *sorry* you were so touchy? or that you hope she will pardon your offence?' (Brontë, *The Tenant of Wildfell Hall*, 1848)

In one token, however, the *that* which prevented the use of an explicit *that*-clause was a demonstrative determiner:

> (31) The despoblado out yonder has a particularly evil name; be on your guard, Caballero. I am *sorry* that Gypsy was permitted to pass; should you meet him and not like his looks, shoot him at once… (Borrow, *The Bible in Spain*, 1842)

This token was surprisingly difficult to interpret due to its ambiguity, but as I studied the original text for more context, it became clear that *Gypsy* was not a proper noun despite the capitalisation. Therefore, *that* is included in the noun phrase as a determiner and the relative *that* has been omitted because of *horror aequi*.

5.2.8 Ø complement

Finally, the zero complement was found to appear in 32 tokens (15.9%). Again, the majority of the tokens in this group were similar to example (32) below:

> (32) He died true to his character – drunk as a lord. Poor lad! I'm *sorry*, too. (Brontë, *Wuthering Heights*, 1847)

However, there were three tokens with the construction *I should be sorry if…,* and of course, one of them was from Anne Brontë (see section 5.2.6):

> (33) Yes,– I have forgiven you: but I know you cannot love me as you once did – and I should be very *sorry* if you were to, for I could not pretend to return it… (Brontë, *The Tenant of Wildfell Hall*, 1848)

A *when*-clause was a popular choice after *sorry* in this sample, too, as there were six tokens which featured this construction.

(34) William and Ready were, therefore, not *sorry* when their work was done, and they went to bed as soon as they had taken their supper. (Marryat, *Masterman Ready*, 1841)

The different senses of *sorry* were spread along this complement group much in the same way as in CLMET 1: the vast majority of the tokens represented sense 2, but there were a few occurrences of senses 1 and 3, too. Example (32) above presents *sorry* in sense 1, while all the other examples in this section thus far have *sorry* in sense 2. Therefore, here is one more example in which *sorry* appears in sense 3.

(35) "Pugh!" said he; "what nonsense have you taken into your brain! Think no more about it. We are all very *sorry*, and beg your pardon; come, shake hands, forgive and forget." (Edgeworth, *The Parent's Assistant*, 1796)

5.2.9 Review of CLMET 2

The number of complementation patterns found in this part of the corpus was a considerable increase from the five patterns in CLMET 1. Altogether eight different patterns were found, the most frequent of which was again the *to*-infinitive with 71 occurrences (35.3%). As was the case in the previous part of the corpus, the construction *should be sorry + to*-infinitive was rather common in this part, too. In addition to the lower clause subject roles of experiencer, agent and theme observed in section 5.1.3, it was found that the subject can also have the role of recipient.

That-clause also maintained its position as the second most frequent complement. Again, the majority of the *that*-clause tokens were of the implicit variety. Changing the subject between the clauses seemed to create a complex enough environment for the explicit *that* to be favoured, because 67% of the explicit tokens had different subjects in the main clause and subordinate clause. Insertions were another factor which seemed to trigger the use of the explicit variety. In the implicit group, there was evidence of the *horror aequi* principle preventing the use of the explicit *that* in at least five tokens.

The non-sentential complements *for* + NP and *at* + NP offered no great surprises. The percentage of the *for* + NP complements was roughly the same as in CLMET 1 (20.7% in the first part and 19.9% in the second), but *at* + NP emerged for the first time, albeit in just two tokens. The gerundial versions, *for* + *–ing* and *at* + *–ing*, were also present with one token each. This could be seen as a tendency towards the Great Complement Shift, as discussed earlier in the theory section.

In addition, the *for* + *wh*-clause pattern appeared in one token, which means that it has undergone quite a substantial decrease from the five tokens encountered in CLMET 1. Finally, there were 32 tokens with no complement, although *if*- and *when*-clauses were again a frequent companion of *sorry*.

5.3 *Sorry* in CLMET 3: 1850-1920

There were 430 occurrences of *sorry* in the third and last part of the CLMET, which extends from the year 1850 to 1920. Again, 50% of those 430 occurrences – that is, 215 tokens – were randomly selected for the following analysis. This data contained 11 tokens in which *sorry* was used attributively:

> (1) "Love is an utterly bygone, *sorry*, worn-out, miserable thing with me – for him or any one else." (Hardy, *Far from the Madding Crowd*, 1874)

As stated above, these tokens are irrelevant regarding the purposes of this study and were therefore discarded. After that, the data consisted of 204 usable tokens, and the complement patterns that were found in this data are presented in the following table.

Pattern	No. of tokens	% of data	NF/million
sorry + *that*-clause	36	17.6	18.08
sorry + *to*-infinitive	49	24	24.61
sorry + *for* + *–ing*-form	2	1	1.00
sorry + *for* + NP	42	20.6	21.09
sorry + *for* + NP$_{REFL}$	1	0.5	0.50
sorry + *about* + NP	2	1	1.00
sorry + Ø	72	35.3	36.16
Total	**204**	**100**	**102.44**

Table 6: Complementation patterns found in CLMET 3.

Interestingly, the *to*-infinitive loses its pole position as the most frequent complement for the first time. Instead, *sorry* + Ø is the pattern with the greatest number of occurrences (72 tokens, or 35.3%). In addition, there were no examples of patterns containing the preposition *at* (*at* + NP or *at* + *–ing*-form) in this part of the CLMET, as there were in the previous part. Other complement constructions, however, will be discussed below.

5.3.1 *For* + NP

The pattern s*orry* + *for* + NP was again the most frequent non-sentential complement with 42 tokens (20.6%). In parts 1 and 2 of the CLMET, the majority of the noun phrases in the *for* + NP tokens were of the type –human and –animate. In CLMET 3, however, 29 of the 42 tokens had a +human and +animate NP in the prepositional phrase, while the remaining 13 tokens represented the –human and –animate type. Below are examples of both types.

> (2) I am very *sorry* for you, Mr. Catchpole, more sorry than I can tell you. (Rutherford, *Catherine Furze*, 1893)
> (3) You make us *sorry* for our manners and habits, if they are so bad; but most of all you are merry at our simplicity. (Meredith, *The Adventures of Harry Richmond*, 1870)

As for the subjects, most of them were again personal pronouns, but there was one token in which the subject was the slightly more unusual *nobody*:

(4) Nobody was very *sorry* for him, for he had been hard upon everyone… (Yonge, *Young Folk's History of England*, 1873)

As was expected, all the tokens where the NP inside the prepositional phrase was +animate took sense 1 of *sorry*. In the –animate NP group, there were occurrences of all senses except sense 4, but this time sense 3 was the most frequent with 6 tokens out of 13 (46%).

Complexity factors were not very common in this sample, as only one token had an insertion in it, and there were no tokens with extractions. The only token to feature insertion can be seen below.

(5) The Buccaneer, watching him go sadly home, felt *sorry* perhaps for his behaviour to the old man. (Galsworthy, *The Man of Property*, 1906)

5.3.2 *For* + NP$_{REFL}$

The *sorry* + *for* + NP$_{REFL}$ pattern appeared in one token in the CLMET 3 sample.

(6) I should be *sorry* for myself if you were to go away; not that I want to put that forward. (Rutherford, *Clara Hopgood*, 1896)

This is the first occurrence of this pattern, which was marked as its own separate subsense in the dictionaries. Here, *sorry* takes sense 4 ('self-pity'). However, as this example is remarkably similar to other tokens containing the construction *I should be sorry if…*, which were interpreted to represent sense 2, the actual difference between these two senses feels unclear. The message of the token would not change drastically if it was paraphrased as follows:

(6') I should be *sorry* if you were to go away; not that I want to put that forward.

Indeed, if we look back into section 4.1, we can see that all the dictionaries combined the *for* + NP$_{REFL}$ pattern with the verb *feel* rather than *be*, which does seem to cause a slight difference in meaning and make the sense more distinct. In example (6) above, therefore, it is unclear whether the author's intention was to denote self-pity or just sadness in general.

5.3.3 *About* + NP

Sorry + *about* + NP emerged for the first time and could be found in two tokens:

> (7) The laundress called, and said she was very *sorry* about the handkerchiefs, and returned ninepence. (Grossmith, *The Diary of a Nobody*, 1894)
> (8) Mr. Vyse is so sensitive. I remember how I used to get on his nerves at Rome. I am very *sorry* about it all, and should not feel easy unless I warned you. (Forster, *A Room with a View*, 1908)

The subjects in both examples take the role of experiencer and the NPs in the prepositional phrases have the semantic features –human and –animate. However, despite the similarities, this pattern cannot be seen as a one-to-one replacement of the absent *at* + NP pattern, because in examples (7) and (8) *sorry* takes sense 3, i.e. apologising for something. The *at* pattern, as we remember from section 5.2.2, featured *sorry* in sense 2. The *for* + NP pattern, on the other hand, seems to be fairly interchangeable with the *about* pattern in cases where apology is meant. Consider, for instance, example (8) above paraphrased with *for*. The meaning remains the same:

> (8') Mr. Vyse is so sensitive. I remember how I used to get on his nerves at Rome. I am very *sorry* for it all, and should not feel easy unless I warned you.

5.3.4 *For* + *–ing-form*

The *for* + *–ing*-form pattern also appeared in two tokens, both of which are presented below.

> (9) "I have left unsaid much that I am sorry I did not say, but I have said little that I am *sorry* for having said, and I am pretty well on the whole, thank you." (Butler, *Note-Books*, 1912)
> (10) Punished again for this, she would own to being *sorry* for saying "won't"; but, bread and water notwithstanding, she never told you. (Collins, *The Moonstone*, 1868)

As was the case with *about*, the role of the subject is experiencer in both cases and *sorry* denotes sense 3. While this is indeed a 100% increase from the one token in CLMET 2, and would seem to suggest that the use of the *for* + *–ing-form* pattern is becoming more frequent, the overall number of tokens with this pattern is still so small that it is rather difficult to make any concrete conclusions regarding The Great Complement Shift theory.

5.3.5 *To*-infinitive

The *to*-infinitive was the second most frequent pattern in this sample with 49 tokens (24%). The

should be sorry + *to*-infinitive pattern observed in the previous sections of the CLMET corpus

seems to have lost its popularity, as only three tokens were found to feature that construction in

CLMET 3.

> (11) … would not the poor old gentleman gladly change lots with me, if he could? I do not
> know; but I should be *sorry* to change lots with him or with any one else, so I need
> not grumble. (Butler, *Note-Books*, 1912)

Extractions were found in two tokens, both of which can be seen below.

> (12) I ain't particular, sir, provided it's sharp, like that poor girl, who, now that she's gone, I
> am *sorry* to have spoke hard on… (Haggard, *She*, 1887)
> (13) I certainly never saw a man, in all my experience, whom I should be so *sorry* to have
> for an enemy. (Collins, *The Woman in White*, 1860)

Insertions, too, were only found in two tokens:

> (14) Just at that moment Barnes appeared, hot with hurrying. 'Very *sorry*, Miss Hopgood, to
> ask you to stay for a few minutes. (Rutherford, *Clara Hopgood*, 1896)
> (15) Perhaps you will be sorry then, *sorry* when it is all too late, to be loved by such as I am.
> (Blackmore, *Lorna Doone*, 1869)

Another interesting feature of example (14) is that the main clause subject (*I*) and the copular verb

(*be*) have been omitted. In fact, there was one other token in the CLMET 3 sample which had a

similar structure, presented below:

> (16) April 27. – Painted the bath red, and was delighted with the result. *Sorry* to say Carrie
> was not, in fact we had a few words about it. (Grossmith, *The Diary of a Nobody*,
> 1894)

Yet even with the subject omitted, it is clear that the implicit subject in examples (14) and (16) is

still *I* and the thematic role is experiencer, as is usually the case with *sorry*. Overall, 41 of the 49

tokens with *to*-infinitive complement in this part of the corpus had *I* as subject, the rest being either

other personal pronouns (*you, she, we*) or proper nouns (*Tom, Miss Clack*).

As for the semantic roles in the sentences containing *to*-infinitives, most of the subjects in

the lower clauses had the role of agent, as was the case in CLMET 2, too. Altogether 31 tokens

were found to represent this type. Again, both sense 2 and 3 of *sorry* were found to appear with

tokens of this kind, as can be seen in the examples below:

> (17) I am *sorry* to say that I have never learnt to read or write. (Collins, *The Woman in White*, 1859–1860)
> (18) "I shall be all right directly. I'm *sorry* to have made a fuss." (Galsworthy, *The Island Pharisees*, 1904)

The thematic role of experiencer, on the other hand, was found in 16 tokens, all of which

took sense 2. Yet again, the verbs *see*, *hear* and *find* accounted for majority of the tokens in this

group, but there were some exceptions, too:

> (19) We are *sorry* to lose him, but perhaps he will get on better by himself… (Grossmith, *The Diary of a Nobody*, 1894)
> (20) Well, I be mortal *sorry* to meet thee in such company… (Hughes, *Tom Brown's Schooldays,* 1857)

In example (20), *meet* could easily be replaced with *see* (or *find*), which is why I interpreted the

subject to have the role of experiencer in this case, too.

In addition, there were two tokens in which the subject took the role of theme. In both of

them, the verb was again passivised, and the sense of *sorry* was sense number 2.

> (21) Sit down, Mr. Catchpole. I am *sorry* to be obliged to impart to you a piece – a something – which is very distressing. (Rutherford, *Catherine Furze*, 1893)
> (22) Perhaps you will be sorry then, *sorry* when it is all too late, to be loved by such as I am. (Blackmore, *Lorna Doone*, 1869)

5.3.6 *That*-clause

There were 36 examples (17.6%) of the pattern *sorry + that*-clause in this data. 26 of them were of

the implicit variety, while the remaining 10 had the *that* written out. As usual, the main clause

subject was commonly *I* or some other personal pronoun, and its thematic role was experiencer. In

the explicit group, sense 2 was the most common sense of *sorry* with six occurrences (60% of all

explicit tokens). Sense 3 was not far behind with four tokens, but there were no occurrences of

sense 1 in the explicit group. The situation was more even in the implicit group, where senses 1, 2

and 3 had 8, 7 and 11 occurrences respectively. Therefore, sense 3 was the most frequent sense in the implicit group, because in 42% of all the implicit tokens, *sorry* took sense 3.

Peculiarly, the number of tokens which had different subjects in the main clause and the lower clause was 6 (60%) in the explicit group and 15 (65%) in the implicit group. Below are examples of both types.

> (23) I feel *sorry* that you should always represent yourself in an unfavourable light.
>   (Gissing, *New Grub Street*, 1891)
> (24) Well, Harry, I'm *sorry* you don't think as we do. (Meredith, *The Adventures of Harry Richmond*, 1870)

This would seem to imply that changing subject between the clauses does not actually create as much complexity as the earlier parts of the CLMET suggested, considering that the number of tokens with different subjects is 5% greater in the implicit group. The CLMET 3 data seems different to the earlier parts also in terms of semantic features of the lower clause subjects: three tokens with −human and −animate subjects were found in the explicit group, but the implicit group was not left far behind with two occurrences of similar tokens. In the earlier parts, the difference between the two groups was much more distinct. Here are examples of both explicit and implicit types with −human and −animate lower clause subject.

> (25) I am *sorry* that my chairs all have wood seats, and are rather hard, but I was thinking of getting some new ones. (Hardy, *Far from the Madding Crowd*, 1874)
> (26) I am *sorry* the house is finished. (Jerome, *They and I*, 1909)

Insertion was only found in one token in the whole sample. In this token, which is presented below, the *that* was explicit. Extractions, on the other hand, were not found to occur with *that*-clause in this part of the corpus.

> (27) He felt half *sorry* at first that he had been confirmed himself. (Hughes, *Tom Brown's Schooldays,* 1857)

However, there was one token in which the explicit use of *that* could be seen as an effort to reduce the complexity factor of the sentence, even though it did not contain any insertions or extractions. Consider the following example (relevant instance of *sorry* italicised):

> (28) …he told people he was sorry he was not more sorry, and here began the first genuine sorrow, for he was really *sorry* that people would not believe he was sorry that he was not more sorry. (Butler, *Note-Books*, 1912)

This thought by Samuel Butler (aptly named *Sorrow within Sorrow)* features extreme repetition of the word *sorry* and multiple *that*-clauses within each other. It can be argued that without the two explicit occurrences of *that*, the whole sentence would be even more difficult to comprehend than it is now.

Finally, there were two tokens in the implicit group which could be seen as having been affected by the *horror aequi* principle because of the proximity of another *that*. Consider:

> (29) 'And do oblige me in a little matter now, Elfride,' said Lord Luxellian warmly, and looking as if he were *sorry* he had brought news that disturbed her. (Hardy, *A Pair of Blue Eyes*, 1873)
> (30) There is going to be such a row in this town that you and he'll be *sorry* you came to it. (Forster, *Where Angels Fear to Tread*, 1905)

In example (29), the *that* which prevented the use of the explicit complement is a relative pronoun and therefore cannot be omitted. In (30), on the other hand, it is a correlative conjunction together with *such* in the higher clause. Leech and Svartvik (2002, 402) mention that when correlative conjunctions are used, "[t]he second marker, if it is *that*, is sometimes omitted". However, because *such a row* and *that* are in this case separated by the adjunct *in this town*, as well as the fact that the subject expression consists of three words instead of one, omitting the second marker would perhaps make the sentence slightly difficult to comprehend (*There is going to be such a row in this town you and he'll be sorry you came to it*). Therefore, it seems to be a better choice to omit the second *that* instead of the first one in order to avoid repetition.

5.3.7 Ø complement

In this part of the corpus, the zero complement was the most common choice and it appeared in more than a third of the tokens. The majority of the Ø complement tokens were, again, variations of the phrase *I'm sorry*, such as the following example:

(31) "Oh, I'm so *sorry*, old chap; I hope I haven't hurt you." (Jerome, *Three Men in a Boat*, 1889)

The role of the subject was experiencer in all tokens and the most common sense of *sorry* was sense 3, although there were some cases in which the sense meant was clearly sympathy rather than apology:

(32) Oh, do you think you must go away before the end of the half? I'm so *sorry*. (Hughes, *Tom Brown's Schooldays,* 1857)

In addition, there were four tokens which featured the familiar *I should be sorry if…* construction, all of which were interpreted to represent sense 2 of *sorry*.

(33) "I should be *sorry*," he said, breaking an awkward silence, "if you were to think me ungrateful… (Galsworthy, *The Island Pharisees*, 1904)

In some cases, both the subject and the verb *be* had been omitted. Altogether, there were five tokens of this kind, two of which are presented below:

(34) "So *sorry*!" my Lady exclaimed, as she and her husband helped him to his feet again. (Carroll, *Sylvie and Bruno*, 1889)
(35) "*Sorry*." "It's no use being sorry if you persist in doing it." (Bennett, *The Old Wives' Tale*, 1908)

All tokens of this kind took sense 3 without exception. Despite the subject being implicit, it is obvious that its thematic role is once again experiencer.

In addition, there was one token which featured a raising construction (discussed in section 2.4.2), as the verb in the main clause was *seem*:

(36) He seemed not so much *sorry*, as scared and dazed like, by what happened. (Collins, *The Woman in White*, 1859–1860)

Thus, this example could be paraphrased as follows:

(36') It seemed that he was not so much *sorry*…

Finally, it must be mentioned that the "adjunct-like" *wh*-clause was again a common companion of *sorry*, as there were three tokens in which it appeared. In all of them, *sorry* represented sense 2.

(37) "Well, Tom, you ain't going to punch my head, I hope, because I insist upon being *sorry* when you got to earth?" (Hughes, *Tom Brown's Schooldays,* 1857)

5.3.8 Review of CLMET 3

Seven different complementation patterns were found in the third part of the CLMET. The most striking difference to the previous part of the corpus is the fact that both complements with the preposition *at* (*at* + NP and *at* + *–ing*), as well as *for* + *wh*-clause, have disappeared. However, there are two new complements, too: *about* + NP and *for* + NP$_{REFL}$.

Furthermore, the *to*-infinitive has lost its position as the most common complement to the zero complement. The distinctive *should be sorry + to*-infinitive construction also seems to have lost its popularity, as only three tokens of this type were found. On the other hand, there was evidence of another, entirely new phenomenon emerging with both the *to*-infinitive and the Ø complement: omitting the subject and the verb altogether.

For the first time, the *for* + NP pattern was more frequent than the *that*-clause complement. A change was also seen in the NPs of the *for* + NP tokens, because unlike CLMET 1 and 2, the majority of them had the semantic features +human and +animate. Therefore, most of them also featured *sorry* in sense 1. *That*-clause complements, too, were different from their counterparts in other sections of the corpus, as the number of tokens which had a different subject in the main and lower clause was – rather unusually – 5% higher in the implicit group.

About + NP and *for* + *–ing* both appeared twice. Despite its similarity to the absent *at* + NP pattern on surface level, the *about* + NP pattern cannot be seen as the former's replacement, because the two pattern occur with different senses of *sorry*. As for the *for* + *–ing* pattern, its 100% increase from CLMET 2 hints at the possibility of The Great Complement Shift in the case of *sorry*, but the overall number of tokens with this pattern is still so small that watertight conclusions are hard to draw.

Finally, there was one occurrence of the *for* + NP$_{REFL}$ pattern. The token in question was remarkably similar to the familiar *should be sorry + to*-infinitive construction, which raised doubts about whether the author actually meant for *sorry* to denote self-pity (as is usually the case with this pattern) or just general sadness/disappointment.

5.4 *Sorry* in the BNC: 1960-1993

It is now time to examine the more recent usage of the word *sorry* by analysing the data that was found in the British National Corpus. Because the material in the CLMET is mostly fictional, the BNC search was restricted to the Imaginative Prose section of the corpus in order to find similar and easily comparable examples. The search string *sorry* returned 4,641 tokens, 10% of which were randomly selected for further analysis. After removing irrelevant results, I was left with 448 usable tokens, the distribution of which can be seen in the table below.

Pattern	No. of tokens	% of data	NF/million
sorry + that-clause	54	12.1	32.73
sorry + to-infinitive	39	8.7	23.64
sorry + for + –ing-form	1	0.2	0.61
sorry + for + wh-clause	1	0.2	0.61
sorry + about + wh-clause	3	0.6	1.82
sorry + for + NP	41	9.2	24.85
sorry + for + NP$_{REFL}$	2	0.4	1.21
sorry + about + NP	19	4.2	11.52
sorry + Ø	288	64.3	174.58
Total	**448**	**100**	**271.57**

Table 7: Complementation patterns found in the BNC.

5.4.1 *For* + NP

41 occurrences of the *for* + NP pattern were found in the data, which makes 9.2% of the whole sample. In 28 tokens, the subject NP was a personal pronoun such as *I* or *she* and the rest were

proper nouns. However, there were two examples in which the personal pronoun was the slightly

more unusual *someone* or *anyone*:

> (1) One day someone is going to feel so *sorry* for her he'll end up marrying her. (HHC 324)
> (2) There was no need for anyone to feel *sorry* for her, Dorothy Fanshawe, of Astbury
> Mews, Upper Grosvenor Street, W.I. (A73 1090)

The NP in the prepositional phrase had the semantic features +human and +animate in 33

tokens, while the remaining eight tokens featured the –human and –animate type. Below are

examples of both types.

> (3) 'If Molly comes home to this mess at Christmas, then I'm *sorry* for her,' asserted Mrs
> Tilling. (ASE 458)
> (4) I am *sorry* for your problems, Emily, and I'll try to help, of course I will, but I have
> worries of my own too. (CKD 2010)

The role of the NP was therefore theme. In all the tokens which had a +human, +animate NP in the

prepositional phrase, *sorry* appeared in the first sense. In the –human and –animate group, too, half

the tokens were found to represent sense 1, but there were another four tokens which were

interpreted to represent sense 3, as it was clear that an apology was implied. An example is

presented below.

> (5) 'Charles Linkworth,' he said at last, 'are you truly *sorry* for your crime?'

Finally, there was one example of an object control construction (discussed in section 2.4.1)

in the *for* + NP complement group. Here, PRO refers to the object of the main clause (*us*) rather

than the subject (*he*).

> (6) "We saw some prisoners being beaten," explained Joseph, turning to Paul Devraux, "but
> he told us not to feel *sorry* for them." (FU8 571)

5.4.2 *For* + NP$_{REFL}$

There were two tokens with the *for* + NP$_{REFL}$ pattern. Their share of the sample was therefore

extremely small, only 0.4%. Both can be seen below.

(7) When I'm being beastly to him, he has such a way of looking *sorry* for himself that I
    begin to hate myself. (G07 3617)
(8) Mr O'Shea feels very *sorry* for himself. (A7J 355)

In both tokens, the role of the subject is experiencer. Both tokens also represent sense 4 of *sorry*, as

is usual with tokens including the *for* + NP$_{REFL}$ pattern. Here, the difference to sense 2 is much

clearer than in the one token of CLMET 3 due to the verbs *look* and *feel*.

5.4.3 *About* + **NP**

The *about* + NP pattern occurred 19 times (4.2%) in this sample, which means that it has overgone

quite a substantial increase from the two tokens (or 1%) in the CLMET 3 data. As usual, the subject

had the semantic qualities +human and +animate and took the role of experiencer. As for the NPs,

however, nouns and pronouns such as *it*, *that* and *this* were again common, but this time there were

also five tokens in which the noun phrases had the semantic features +human and +animate. The

tokens in CLMET 3 only had –human and –animate NPs. Regardless of the type, the role of the NP

was theme. Here are examples of both human and non-human NPs:

(9) 'I'm *sorry* about Zadak,' he said. (FSL 671)
(10) 'All right, I'm *sorry* about the knife in the pond,' he said, raising his hands in
     submission. (GWG 1936)

As usual, *sorry* took sense 1 if the NP in the prepositional phrase was +human and +animate. If the

NP did not have these features, *sorry* took either sense 1 or sense 3.

Finally, in five out of the 19 *about* + NP tokens (26%), both the subject and the verb had

been omitted. Here are two examples:

(11) '*Sorry* about Peter.' (CDN 547)
(12) '*Sorry* about all this,' he said, 'my name's Hennessy.' (G03 377)

5.4.4 *For* + *–ing-form*

There was one example of *for* + *–ing*-form in this data.

(13) I say I'm *sorry* for treading on her, but she just looks at me like she don't understand. (A74 2957)

As usual, the role of the main clause subject is experiencer, while the lower clause subject takes the role of agent. Judging by the context, it is quite obvious that an apology is meant, therefore *sorry* in this token represents sense 3.

Again, it is rather difficult to say whether the –*ing*-form was chosen because of The Great Complement Shift (as this sentence could be paraphrased – albeit rather formally – as *I'm sorry to have tread on her*) or whether the –*ing*-form has replaced a *that*-clause because of *horror aequi* (*I say [that] I'm sorry that I tread on her*). In any case, the fact that there is still only one occurrence of this pattern does not shed any more light on whether the gerund is becoming more frequent. However, at least one conclusion can be made: it has come to stay.

5.4.5 *For* + *wh*-clause

For + *wh*-clause also appeared in only one token, which is presented below.

(14) Always before, after similar wickednesses, she had felt almost gleeful triumph, but this time, although she was not truly *sorry* for what she had done to Mrs Darrell, she felt something like remorse. (HGE 959)

In this case, too, *sorry* seems to represent sense 3 and the lower clause verb is again one of the three options observed earlier: *do*, *say* and *happen*. The subject *she* has the role of experiencer in the main clause, but takes the role of agent in the subordinate clause.

5.4.6 *About* + *wh*-clause

About appeared together with a *wh*-clause, too. This was the case in altogether three tokens, all of which can be found below.

(15) Perhaps it was this, the fact that she now had someone else to love her, that made it possible for me to make a proper apology, to stutter out confusedly, red-faced, that I was *sorry* about what happened. (FR3 1718)

(16) I just… . it's just that I wanted to tell her how *sorry* I am about … about what happened to your father and brother but, not knowing her, I didn't want to say it in a way that might upset her. (FSC 1544)

(17) Look, I was *sorry* about what happened round at the flat. (G15 2124)

In all these tokens, the lower clause verb is *happen*, which leaves the subject of the clause unclear. This could be a conscious choice in example (15), where it is clear that the protagonist has done something he wants to apologise for, but perhaps does not want to spell it out by using a first person pronoun that would immediately label him as the guilty party. Instead, he uses the ambigous term 'happen'. In example (16), it seems more likely that the speaker is just feeling sympathetic and trying to express his or her condolences, which means that *sorry* represents sense 1 rather than sense 3. Example (17), however, could be interpreted to denote either sense 1 or sense 3, but a look into the BNC confirms that sense 3 is the correct one:

> $_{2123}$ 'About wanting to meet you. $_{2124}$ Look, I was sorry about what happened round at the flat. $_{2125}$ It was… well.' $_{2126}$ The detective waited to hear whether Roger would apologise. $_{2127}$ He never apologised. (G15 2124-2127)

Here, too, the detective is expecting Roger to make a proper, personal apology instead of vaguely referring to something that just 'happened'. It seems, therefore, that if the verb *happen* is followed by the preposition *to* and a NP (as in example 16), we can expect *sorry* to take sense 1, but if it occurs on its own, a somewhat veiled apology is implied.

5.4.7 *To*-infinitive

The *to*-infinitive complement accounted for 39 tokens (8.7%). The *should be sorry* + *to*-infinitive construction, which was quite common in all parts of the CLMET, seems to have now disappeared completely, as no tokens of this kind were found in the BNC. Omitting the subject and verb of the main clause, on the other hand, has become increasingly frequent. Altogether 12 tokens of this kind were found (31%), two of which can be seen below.

(18) '*Sorry* to take so long!' (CEB 541)

(19) Andrew, so *sorry* to interrupt you but I wanted to say hello. (CS4 1647)

No extractions or insertions were found in this sample, nor was there any evidence of the *horror aequi* principle. In general, the sentences containing the *to*-infinitive construction were rather short (like example 18 above), which might be the reason why no complexity factors were found.

While the role of the main clause subject was again experiencer in all tokens, the lower clause subjects were found to have the role of either experiencer or agent. The agent role was the more common of the two with 29 tokens out of 39 representing this type. Out of the 29 agent-type tokens, 25 had sense 3 of *sorry*, while the remaining four tokens took sense 2. Below are examples of both.

(20) '*Sorry* to butt in.' (FNU 22)
(21) 'You know, in some ways I'm most frightfully *sorry* to leave.' (BMU 80)

In all the tokens which had the role of experiencer in the lower clause, however, *sorry* was interpreted to mean sense 2. There were 10 tokens of this kind, two of which are presented below.

(22) Tell her I'm not angry, just *sorry* to have lost her. (GWH 1035)
(23) So I was not *sorry* to see them go. (H9N 1240)

5.4.8 *That*-clause

The second largest complement group in this data was the *that*-clause with 54 occurrences (12.1%). Again, these were divided into two groups according to whether the *that* was explicit or implicit. The implicit variety was considerably more frequent with 47 tokens, while the explicit type only appeared in seven tokens. The subject of the main clause was commonly *I* or some other +human and +animate NP, and took the role of experiencer. However, there was one rather unusual exception:

(24) 'I'm *sorry* I pounced on you before,' the dog told them. (CFJ 1480)

Here we have a rare case in which it is clear that *I* does not refer to a +human subject as one might normally presume. In real life, animals obviously cannot talk, at least not in a way understandable to humans, which is why they do not refer to themselves with the pronoun *I*. However, as this

section of the corpus is called Imaginative Prose, –human subjects are perfectly acceptable, too.

Sense 3 was the most prevalent sense of *sorry* in this complement group. There were some occurrences of senses 1 and 2 in both explicit and implicit groups, but as their percentages were roughly the same in both groups, it is not likely to be a factor that affects complement choice. Below are examples of senses 1 and 2.

(25) I'm so *sorry* you're the devil-man's daughter. (HTN 1175)
(26) If you're *sorry* you married me, well, so am I! (FRE 1631)

As was observed earlier in this section with other complement types, omitting the main clause subject and verb seems to have increased. There was evidence of this phenomenon in the *that*-clause data, too, as seven tokens (13%) of this kind were found. All of them represented the implicit group.

(27) *Sorry* I haven't been able to get Sergeant Henley for you. (H85 3708)

As for the lower clause subjects, six out of seven tokens (86%) in the explicit group had different subjects in the main clause and lower clause. The lower clause subject had the semantic features –human and –animate in three of them, one of which can be seen here:

(28) They'd come up together in the crime world, she reflected and she was *sorry* that her findings were clearly causing him trouble. (HA2 1195)

In the implicit group, on the other hand, only 18 of the 47 tokens (38%) had different subjects in each clause, and of these, only two were –human and –animate. Again, an example of this type of token can be seen below.

(29) If this is my last word on the subject, I'm *sorry* it has to be so close to bathos, but that's how I feel. (G0N 390)

Therefore, even though the evidence in CLMET 3 appeared to suggest the contrary, changing the subjects between clauses does seem to have an effect on whether the explicit or implicit variety of *that* is chosen.

Two occurences of insertion were found in the data. Both of them appeared with tokens

belonging to the explicit group, which implies that in the case of the *that*-clause complement in the

BNC, the complexity principle seems to hold true. Both tokens containing an insertion are presented

below.

> (30) I am *sorry*, James, that this man was your father. (FRD 346)
> (31) I'm *sorry*, in a way, that my unorthodox matchmaking exercise has come to nothing.
>   (HGN 3645)

5.4.9 Ø complement

As can be seen from the table at the beginning of this section, the vast majority (288, or 64.3 per

cent) of tokens in the BNC were of the zero complement type. In fact, 26 of them consisted of

nothing but the word *sorry*, and 35 tokens only contained the words *I'm sorry*, as in the following

examples:

> (32) '*Sorry*.' (BMR 1379)
> (33) 'I'm *sorry*.' (CDE 675)

In addition to one-word tokens such as example (32) above, there were also longer tokens in which

the subject and the verb had been omitted. All in all, there were 87 tokens of this kind, which makes

30 per cent of the whole Ø complement data. Examples can be seen below.

> (34) '*Sorry* – didn't mean to be nosy.' (JYC 646)
> (35) She told him *sorry*. (FAS 822)

> Sense 3 was by far the most common sense of *sorry* to occur with the zero complement, but

there were exceptions, too. 28 tokens (10%) were interpreted to represent sense 1:

> (36) 'I'm so *sorry*, Delia,' Rosen said, and immediately thought: Blast, remembering her
>   cold reception of his sympathy on the telephone the night before. (FPF 67)

There were also 10 tokens (3%) in which *sorry* took sense 2. In this group, the construction *will be*

sorry, meaning 'regret', was particularly common:

> (37) 'Sit down and write that cheque,' he commanded, 'or you'll be bloody *sorry*.' (FRJ
>   1175)

However, there were no tokens with the closely related construction *should be sorry if…*, which was

quite frequent in the CLMET.

In addition, there was one token in which I interpreted *sorry* to represent sense 4, even though it was indeed a Ø complement token and did not feature the distinctive *for* + NP$_{REFL}$ pattern. While I was comparing my data to the original texts in the BNC corpus in order to gain more context for the numerous 'I'm sorry' –type tokens and thus determine the sense that was being used in each token, I came across this one:

(38) I'm not *sorry*. (FSP 205)

At first glance, this may seem to be another token where *sorry* takes sense 3, but as I referred to the corpus to confirm my initial assumption, it became clear that the token in question was, in fact, an example of sense 4 in disguise. Consider:

203 'Anyway, I'm too fat to get married.'
204 'Don't be so sorry for yourself.'
205 'I'm not sorry. 206 I'm just fat.' (FSP 203-206)

Therefore, it could be said that the complement choice (or the lack thereof) in this token was affected by the desire not to repeat the *for* + NP$_{REFL}$ pattern, i.e. the *horror aequi* principle.

5.4.10 Review of the BNC

As many as nine different complementation patterns were found in the BNC data, which is more than in any part of the CLMET. Again, the zero complement was the most common choice, accounting for 64.3 per cent of the whole BNC data. In 30% of the Ø complement tokens, both the subject and the verb had been omitted. That was a frequent feature also in the second most common complement group, the *that*-clause, in which 13% of the tokens were of this type.

Based on the *that*-clause complements in the BNC data, it seems that CLMET 3's higher percentage of different main and lower clause subjects in the implicit group was somewhat of an anomaly, because in this corpus, most tokens with different subjects were again found in the explicit group. Insertions were also found to add complexity and require the use of the explicit *that*.

For + NP was the third most common pattern with 41 occurrences, meaning that the *to-*infinitive, which was the most common pattern in both CLMET 1 and 2, was left at fourth place in the BNC. The majority of the *for* + NP tokens had a +human NP in the prepositional phrase and therefore featured *sorry* in the first sense. In the *to*-infinitive group, on the other hand, it was found that the *should* construction seems to have now disappeared completely, while the omitting the subject and the verb altogether has increased.

The *about* patterns, *about* + NP and *about* + *wh*-clause, have also become increasingly frequent. The former occurred 19 times, while the sentential variety was found in three tokens. For the first time, the NP in the *about* + NP construction could also have the semantic features +human and +animate, although the inanimate variety was still more common. There was evidence in the *wh*-clause group that if the verb *happen* is followed by the preposition *to* and a NP, *sorry* takes sense 1, but if it occurs on its own, sense 3 is implied.

Finally, there were a handful of tokens containing different patterns with the preposition *for*: *for* + NP$_{REFL}$, *for* + *–ing* and *for* + *wh*-clause. This time, it was more evident that sense 4 was indeed meant in the tokens containing the *for* + NP$_{REFL}$ pattern. The other two patterns offered nothing revelatory aside from the re-emergence of the *for* + *wh*-clause pattern after its absence in CLMET 3. However, the share of both *for* + *wh*-clause and *for* + *–ing*-form in the whole data was still infinitesimal.

6 Summary and concluding remarks

In this thesis, I have analysed 993 tokens from two corpora of British English in which the adjective *sorry* was used predicatively. Altogether 11 different complementation patterns were found to have been used with *sorry* during the 283 years covered by the CLMET and the BNC corpora. My aim with this thesis, besides mapping all the different patterns that can be used with this adjective, was to find out how the complementation of *sorry* has changed during the last three centuries and whether there are any patterns which have been frequent at one time but fallen into disuse later on. I also wanted to explore the meanings attached to these patterns and discover possible changes in the ways different patterns are used.

The table below is a summary of all the different patterns found in each time period.

Time-frame	Number of types	Complement types
CLMET 1: 1710–1780	Sentential: **3**	*To*-infinitive; *that*-clause; *for* + *wh*-clause
	Non-sentential: **1**	*For* + NP
	Zero complement	Ø
CLMET 2: 1780–1850	Sentential: **5**	*To*-infinitive; *that*-clause; *for* + *wh*-clause; *for* + *–ing*; *at* + *–ing*
	Non-sentential: **2**	*For* + NP; *at* + NP
	Zero complement	Ø
CLMET 3: 1850–1920	Sentential: **3**	*To*-infinitive; *that*-clause; *for* + *–ing*
	Non-sentential: **3**	*For* + NP; *for* + NP_{REFL}; *about* + NP
	Zero complement	Ø
BNC: 1960–1993	Sentential: **5**	*To*-infinitive; *that*-clause; *for* + *wh*-clause; *about* + *wh*-clause; *for* + *–ing*
	Non-sentential: **3**	*For* + NP; *for* + NP_{REFL}; *about* + NP
	Zero complement	Ø

Table 8: Summary of the patterns found in the CLMET and the BNC.

The highest number of patterns was found in the BNC, which presented us with a total of nine different complements: five sentential and three non-sentential complements, plus the zero complement. Conversely, the lowest number of patterns appeared in the first part of the CLMET, where only four different complements plus the zero complement were found. The increase of the

patterns has not been linear, however, because from 1780 onwards, the number decreases from eight in CLMET 2 to seven in CLMET 3 and then goes back up to nine in the BNC. On the other hand, it can be safely assumed that these numbers are far from a perfect description of actual usage: the *for + wh*-clause pattern, for example, is extremely likely to have been used also between 1850 and 1920 despite not being present in CLMET 3, because it appears in both CLMET 2 and the BNC. It seems, therefore, that the number of complements has stayed roughly the same from the 1800s onwards.

However, the complement types have not remained the same, because some patterns have indeed disappeared while new patterns have emerged. The two patterns containing the preposition *at*, *at + NP* and *at + –ing*, only appeared in the second part of the CLMET and have not been seen since. On the other hand, the use of *about* (with both NP and *wh*-clause) has increased towards the 21st century. Below is a figure which sheds more light on the development of the six most frequent patterns during the three centuries.
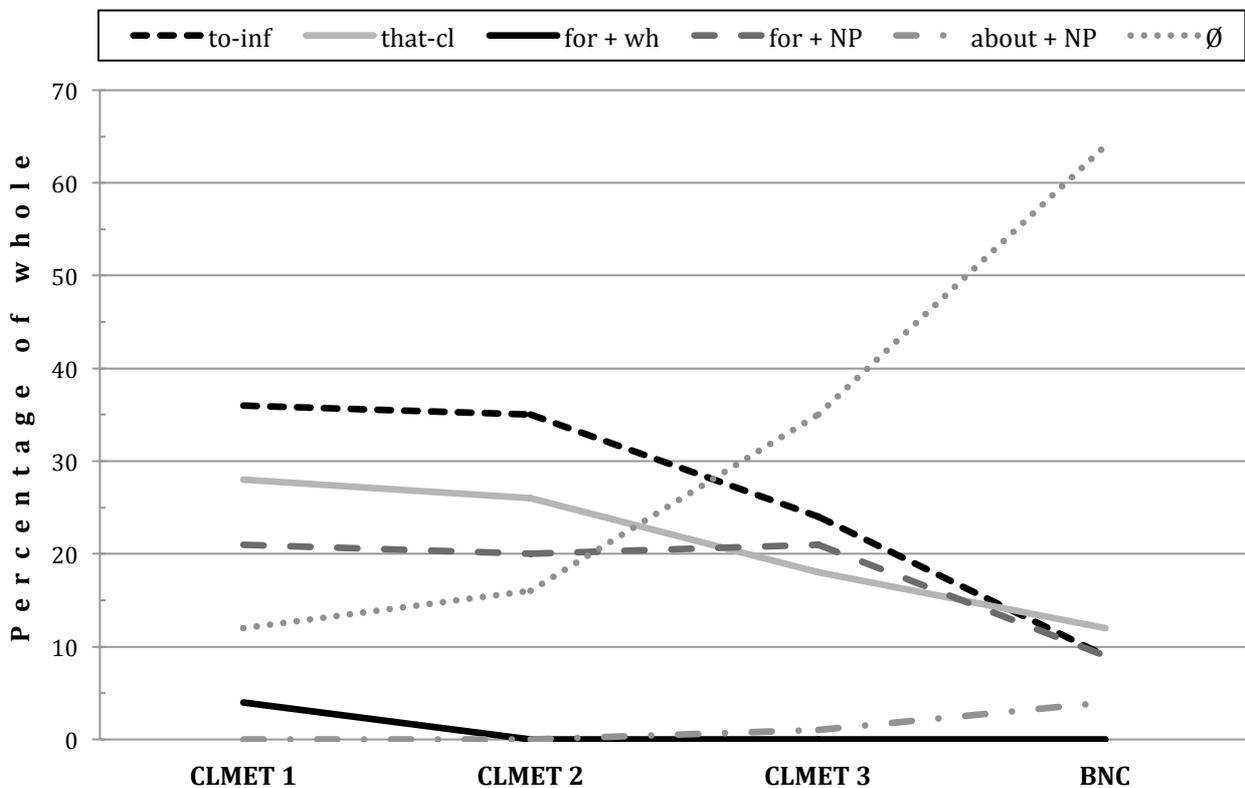

Figure 2: The development of the six most frequent complementation patterns.

Here we can clearly see that even though the frequency of all the sentential patterns is generally decreasing, the *that*-clause has become slightly more frequent than the *to*-infinitive. In addition, the *for* + NP pattern has suffered a steep decline between CLMET 3 and the BNC and seems to be giving way to the *about* variant. However, the most striking development of all is the fact that *sorry* is increasingly used on its own without any complement. The share of the zero complement of the whole corpus data has grown from meagre 12.1% in CLMET 1 to an overwhelming 64.3% majority in the BNC.

   In relation to the zero complement discovery mentioned above, it was also found that *sorry* increasingly occurs without any subject nor verb, regardless of the complement type. At the same time, the frequency of the *should be sorry* construction has declined. Below is another figure describing this development.
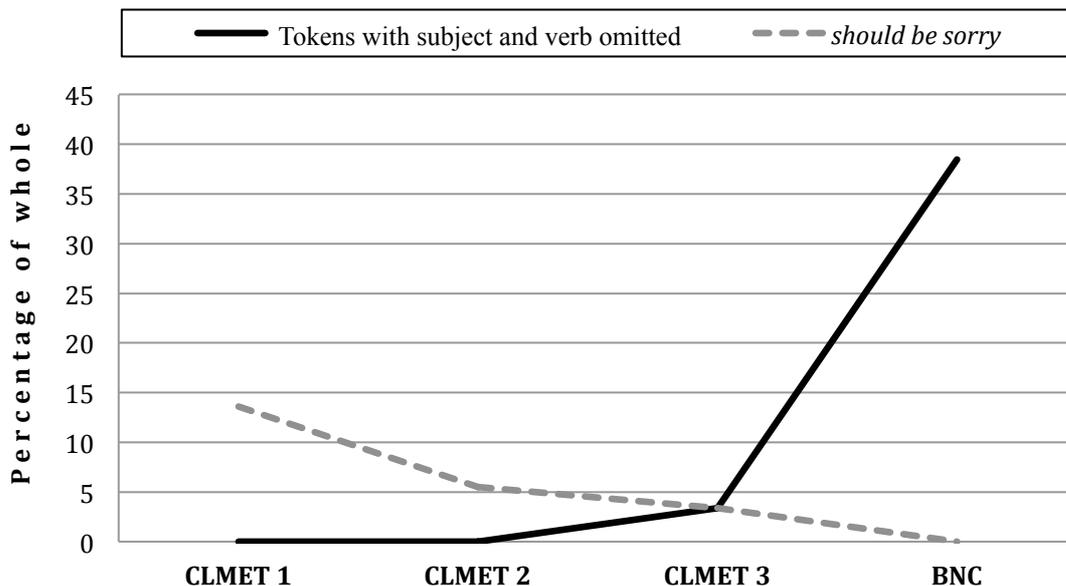

Figure 3: The percentages of tokens with no subject nor verb and tokens with the should be sorry *construction.*

It seems, therefore, that the usage of *sorry* is becoming more and more simplified and economical: why use complex complementation patterns and subject–verb constructions, when in most cases, simply the word *sorry* will do? Based on the data, it could even be argued that *sorry* seems to be

losing its adjectival qualities and becoming more like a one-word exclamation. This is also supported by the changes concerning the different senses of *sorry*, which are discussed more thoroughly in the next paragraph.

Here is a figure presenting all the senses of *sorry* encountered in different parts of the CLMET and in the BNC.
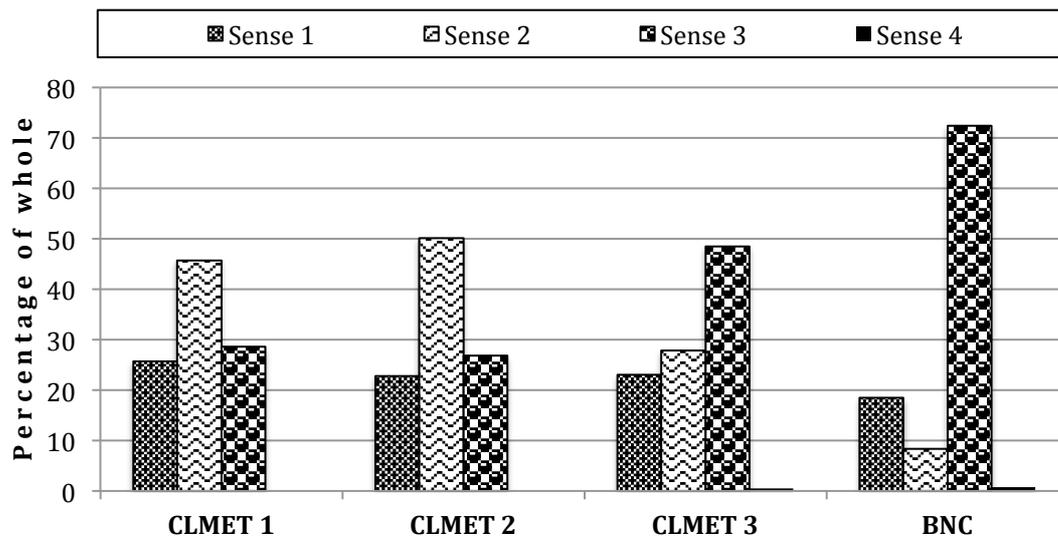

Figure 4: The percentages of the different senses in each period.

The second sense ('sadness, disappointment or regret') was the most common sense of *sorry* in both CLMET 1 and 2, but from CLMET 3 onwards, the majority of the usage of *sorry* has focused on the third sense, i.e. 'apology'. In the BNC, sense 2 has in fact even fallen behind the first sense ('sympathy') in frequency. Therefore, as was observed in the previous paragraph, the use of *sorry* has shifted towards the sense in which simply the word itself is enough and no explanatory complementation patterns are needed – although it must be noted that *sorry* can also occur on its own without complements in senses 1 and 2, as was found out earlier in the corpus analysis. However, sense 3 is still the most common sense to appear with the zero complement, at least in modern usage: in the BNC, 250 out of the 288 tokens with no complement (87%) featured *sorry* in sense 3.

A shift in meaning can also be seen in the *about* + NP pattern. When it first appeared in CLMET 3, all its tokens featured *sorry* in sense 3. It could therefore be seen as a more distinct alternative to denote sense 3 instead of the *for* + NP pattern which can occur together with senses 1, 2 and 3 and the meaning of which may therefore be more difficult to distinguish. In the BNC, however, it emerged that the usage of the *about* pattern has spread to include sense 1, too, as evidence of +human NPs was found. At the same time, the frequency of the *for* + NP pattern has declined. It remains to be seen whether *about* + NP will at some point overtake *for* + NP altogether or whether both constructions continue to exist side by side.

Overall, *for* +NP, *that*-clause and the zero complement were the least selective patterns concerning the different senses, as all of them were found to occur with all the senses except sense 4. In fact, the zero complement was somewhat unusually found even with sense 4 ('self-pity') in the BNC. The most selective patterns, according to my data, were *for* + NP$_{REFL}$ (occurred exclusively with sense 4), *for* + *wh*-clause (sense 3), *for* + *–ing* (sense 3), and the two patterns with *at* (sense 2). *About* + NP and *about* + *wh*-clause both occurred with senses 1 and 3, while the *to*-infinitive could be found with *sorry* meaning either sense 2 or 3.

With the *to*-infinitive, it was observed that when the subject of the lower clause has the thematic role of experiencer or theme, the sense of *sorry* is always sense number 2. If the subject takes the role of agent, the sense can be either 2 or 3. *For* + NP was also found to have similar restrictions concerning the senses: if the NP has the semantic features +human and/or +animate, *sorry* is most likely to take sense 1. However, if it does not, *sorry* can appear in all senses except sense 4. Finally, the BNC data suggested that if the verb *happen* is used in the *wh*-clause of an *about* + *wh*-clause pattern, and followed by the preposition *to* and a NP (e.g. *I'm sorry about what happened to your father*), *sorry* can be expected to take sense 1. If *happen* occurs on its own without the prepositional phrase, however, sense 3 is strongly implied. It can therefore be concluded that at least to some extent, the sense which is meant affects the structure that is chosen, and vice

versa.

As for the various principles affecting complement choice that were introduced in the theory section of this thesis, it was found that the complexity principle does have an effect on whether the explicit or implicit variety of *that* is chosen when the *that*-clause complement is used with *sorry*. Having different subjects in the main and lower clause, as well as insertions and extractions, generally seem to add complexity and thus trigger the use of the explicit *that* – despite some fluctuation witnessed in CLMET 3. There was also evidence of *that* being omitted because of the *horror aequi* effect.

The frequency of gerundial complements (*for + –ing* and *at + –ing*) was incredibly small in this data, which – as stated earlier – makes it rather difficult to make any definite conclusions concerning The Great Complement Shift in the case of *sorry*. *At + –ing* only appeared in CLMET 2, which points towards the verdict that it is nowadays quite an archaic choice of complement. *For + –ing*, on the other hand, was present in both CLMET 2 and 3, as well as the BNC. It can be said, therefore, that even though its frequency was only 0.2–1% depending on the time period, and it seems unlikely to overtake the *to*-infinitive any time soon, the gerundial complement (especially combined with *for*) is a perfectly acceptable choice to accompany *sorry* and seems to have stabilised its position. Whether this apparent reluctancy to favour the gerund over the infinitive is a universal phenomenon or an exclusive feature of British English (as was discussed in section 2.5.4) is an interesting question which may perhaps even be worthy of further research.

Finally, the following figure displays the overall frequency of *sorry* in its predicative sense throughout the three centuries discussed in this thesis.
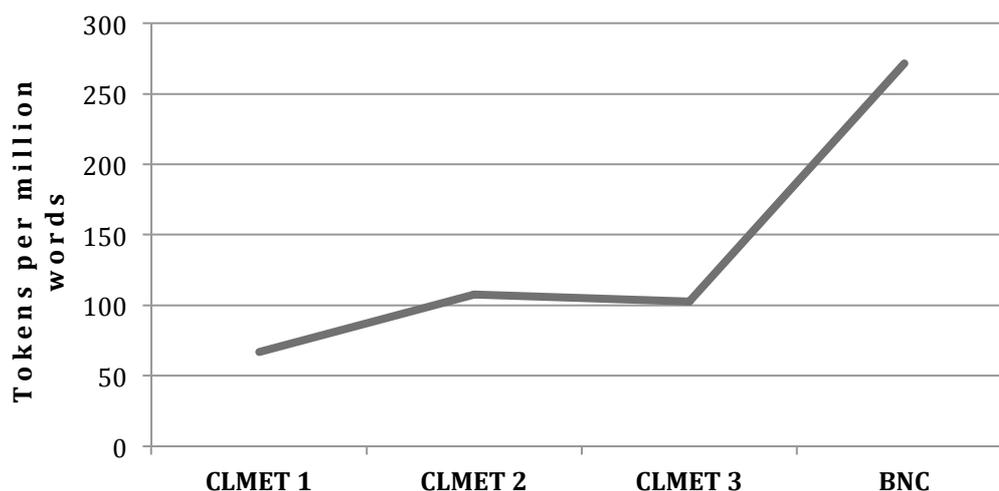
Figure 5: The development of the normalised frequency of sorry

The overall frequency of tokens with the word *sorry* used predicatively experienced a minor drop between the second and third part of the CLMET, but for the most part, the trend has been ascending. In the BNC, the frequency is the highest it has ever been, which is no doubt connected to the increasingly common zero complement and sense 3 becoming the most prevalent sense of *sorry*, both of which were discussed above. It can therefore be expected that this growing trend will continue also in the future, unless there are some major changes in the use of British English.

In this thesis, I have examined the different aspects of the behaviour of *sorry* that I stated as particular points of interest in the introduction. I have achieved the goals I set for myself and found some interesting results along the way. However, there were some questions which were left unanswered due to the limited scope of the present thesis. For example, further research into the gerundial complements of *sorry* might prove fruitful and interesting, as there was very little evidence of the *–ing*-form in this data.

References

Primary sources

CLMET – *The Corpus of Late Modern English Texts*

BNC – *The British National Corpus*

Secondary sources

Biber, Douglas and Susan Conrad, Randi Reppen. 1998. *Corpus Linguistics.* Cambridge: Cambridge University Press.

Biber, Douglas et al. 1999. *Longman Grammar of Spoken and Written English.* London: Longman.

Burnard, Lou (ed.) 2007. *Reference Guide for the British National Corpus (XML Edition).* Oxford, Oxford University. Available from <http://www.natcorp.ox.ac.uk/docs/URG/> [accessed April 2013]

Carnie, Andrew. 2007. *Syntax: A Generative Introduction*. 2nd ed. Oxford: Blackwell.

De Smet, Hendrik. 2005. "A Corpus of Late Modern English Texts." *ICAME Journal* 29, 69-82.

Dowty, David. 1991. "Thematic Proto-Roles and Argument Selection." *Language* 67, 3: 547-619.

Fanego, Teresa. 1996a. "The Development of Gerunds as Objects of Subject-control Verbs in English (1400-1760)." *Diachronica* 13, 29-62.

Fanego, Teresa. 2007. "Drift and the Development of Sentential Complements in British and American English from 1700 to the Present Day." In *Of Varying Language and Opposing Creed: New Insights into Late Modern English*, eds. Javier Pérez-Guerra, Dolores González-Álvarez, Jorge L. Bueno-Alonso, Esperanza Rama-Martínez, 161-235. Bern: Peter Lang.

Haegeman, Liliane. 1991. *Introducing Government and Binding Theory*. Oxford: Blackwell.

Herbst, Thomas et al. 2004. *A Valency Dictionary of English*. Berlin: Mouton de Gruyter.

Huddleston, Rodney and Geoffrey Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Lakoff, George and Ross, John R. 1966. *Criterion for Verb Phrase Constituency. In Harvard Computation Laboratory Report to the National Science Foundation on Mathematical Linguistics and Automatic Translation*, No. NSF-17. Cambridge, MA: Harvard University Computation Laboratory.

Leech, Geoffrey and Jan Svartvik. 2002. *A Communicative Grammar of English.* London: Longman.

The Oxford English Dictionary. 3rd ed. 2009. [Internet] Oxford: Oxford University Press. Available from http://www.oed.com. [Accessed October 2012 – January 2013]

Postal, Paul M. 1974. *On Raising: One Rule of English Grammar and its Theoretical Implications.* Cambridge: MIT Press.

Postal, Paul M. 1994. "Contrasting Extraction Types." *Journal of Linguistics* 30, 159-186.

Quirk, Randolph and Sidney Greenbaum, Geoffrey Leech, Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Radford, Andrew. 1988. *Transformational Grammar: A First Course*. Cambridge: Cambridge University Press.

Rohdenburg, Günter. 1996. "Cognitive complexity and increased grammatical explicitness in English." *Cognitive Linguistics* 7, 2: 149-182.

Rohdenburg, Günter. 2003. "Cognitive complexity and *horror aequi* as factors determining the use of interrogative clause linkers in English." In *Topics of English Linguistics 43: Determinants of Grammatical Variations in English*, ed. Günter Rohdenburg and Britta Mondorf, 235-237. Berlin: Mouton de Gruyter.

Rohdenburg, Günter. 2006. "The Role of Functional Constraints in the Evolution of the English Complementation System." In *Syntax, Style and Grammatical Norms*, ed. Christiane Dalton-Puffer, Dieter Kastovsky, Nikolaus Ritt and Herbert Schendl, 143-165. Bern: Peter Lang.

Rudanko, Juhani. 2006. "Watching English Grammar Change: A Case Study on Complement Selection in British and American English." *English Language and Linguistics* 10.1: 31-48.

Sinclair, John, ed. 2003. *Collins COBUILD Advanced Learner's Dictionary.* 4th ed. Glasgow: HarperCollins Publishers.

Somers, Harold. 1984. "On the Validity of the Complement-Adjunct Distinction in Valency Grammar." *Linguistics* 22, 507-530.

Vosberg, Uwe. 2003a. "The role of extractions and *horror aequi* in the evolution of *–ing* complements in Modern English." In *Topics of English Linguistics 43: Determinants of Grammatical Variations in English*, ed. Günter Rohdenburg and Britta Mondorf, 305-327. Berlin: Mouton de Gruyter.

Vosberg, Uwe. 2003b. "Cognitive Complexity and the Establishment of *–ing* Constructions with Retrospective Verbs in Modern English." In *Linguistic Insights: Insights into Late Modern English*, ed. Marina Dossena and Charles Jones, 197-220. Bern: Peter Lang.

Vosberg, Uwe. 2009. "Non-Finite Complements." In *One Language, Two Grammars? Differences Between British and American English*, ed. Günter Rohdenburg and Julia Schlüter. Cambridge: Cambridge University Press.

Wehmeier, Sally, ed. 2005. *Oxford Advanced Learner's Dictionary*. 7th ed. New York: Oxford University Press.