KU LEUVEN
FACULTY OF ARTS
DEPARTMENT OF LINGUISTICS

# English phrases, French verbs

Causes and consequences of
loan word accommodation biases

Thesis presented in partial fulfilment of the requirements
for the degree of Doctor in Linguistics

**Marlieke Shaw**

Supervisor:

Prof. Dr. Hendrik De Smet

Leuven, 14 December 2022

There are a whole lot of things in this world of ours
you haven't started wondering about yet.

*Roald Dahl*

# Contents

# List of Tables

X

# List of Figures

# Acknowledgements

Starting PhD research is like receiving a parcel of land that needs to be cultivated. Like in horticulture, one cultivates the land one day at a time: every letter, conversation, or analysis contributes to making the garden bloom. At times, the garden only seems to wilt, and other times it goes from winter to summer in a single day. Except, in horticulture no one would intrude in the neighbour's garden to voluntarily wield it, replace flowers by prettier flowers, or even sow new plants where the neighbour has left a bare patch. In a way, research is more collaboratory than horticulture, because other parcels are just as important as one's own. However, sometimes the sky is overcast, the flowers look dull, and nothing seems to grow. A turbulent undertaking like a PhD can only be done with help.

I could not have undertaken this journey without my supervisor, Hendrik De Smet, who has offered me a parcel of land and who has patiently taught me how to dig it. Thank you for our intriguing conversations, your ability to put things into perspective, for generously sharing your insights, and for doing all that with an admirable mix of consideration and determination. I also want to thank Lies Sercu for convincing me four years ago that I am capable of growing and tending to a garden plot.

I want to express my sincere gratitude to Michèle Goyens, Bettelou Los, Carola Trips, and Eline Zenner for agreeing on evaluating my garden as members of my examination committee, and Toon Van Hal for acting as chair of this jury. I owe special thanks to Michèle Goyens and Carola Trips, who have also been faithful members of my advisory committee.

I am very grateful to Carola Trips for selflessly welcoming me in her garden. Her weekly colloquia have been a safe space for discussing my progress on a regular basis. The Anglistik IV research group has always been welcoming, encouraging, and generous with their feedback. I want to thank all of the research members, but in particular Juliane Elter, Lena Kaltenbach, Mareike Keller, Michael Percillier, and Nadine Zürn for their critical yet constructive comments along the way. My garden has greatly benefited from your cross-pollination.

Marie-Anne, you have been my partner in crime during this horticulture bootcamp. You have taken on the thankless role of reminding me of my limits on moments when I was too stubborn to respect them myself. There are so many things I would like to thank you for: for proofreading my thesis, but also for all our chats and laughter, for your help when I was messing with statistics, for figuring out things together, your baking skills, and that you tried that deep-fried Mars bar with me.

I am grateful to the researchers who have started cultivating their gardens around the same time as I have: Afra, Alexandra, Catho, Charlotte, Ellison, Giulia, Jorina, and Virginia. I also want to thank my other colleagues, many of whom have become friends: Atila, Céline, Federica, Katrien (Arron), Lena (Heynen), Manon, Melissa, Miriam, Paloma, and all the others. Special thanks go to Jorina, for our trips together, for your emergency proofreading, your exceptional thoughtfulness and optimism, our goofy messages, for providing cocktail nuts in times of distress, and for always keeping on going. Ellison, thank you for our culinary discoveries, trips, musical intermezzos, your infectious peacefulness, your energy, and your curiosity for life. Thank you, Virginia, for adopting me as your little sister, for feeding me *pistacchio*, for your wise words, and your hugs when there are no words. Giulia, thank you for always believing in me, and for going the extra mile together. Dana, your enthusiasm, positive vibes, and advice have always been cathartic.

I wish to thank my friends and family for their soothing sunrays, and for dealing with my distracted mind after having spent too many long nights in a row in my garden. Atila and Lars, thank you for including me in your weekend plans. Indya, thank you for being such a good friend that you even make your handwriting legible for me. Adrianna and Céline, I am proud of our "club des femmes fortes". Lindsay, thank you for our chats, walks, and your recuperative afternoon pastries. Juliane, thank you for our smooth collaboration, your never-ending support, and our beautiful moments in Oxford. Iris, I can feel you being grounded at the other side of the Canal. Thank you for those recharging days at the English coast, our long walks, virtual movie nights, your never-waning belief in me, and for always making me laugh.

Special thanks go to Pedro, for proofreading the translations of my examples, for his patience, encouraging words (even when I did not want to listen), hugs, delicious meals, unconditional love, never-lacking humour, and *joie de vivre*. Thank you for getting me out of the garden when it was raining too hard.

Last, thank you — reader — for visiting my garden.

# Chapter 1

# Introduction

## 1.1. Loan word accommodation and its constraints

"When a foreign word falls by accident into the fountain of a language, it will get driven around in there until it takes on that language's colour" (Jakob Grimm in Campbell 1998: 57). In less colourful words than Grimm's, the phenomenon in which foreign words (or *loan words*) fall into the fountain of a language is also known as *borrowing*. In this process, the language which provides the foreign material is called the *source language*, and the language which adopts the material is called the *recipient language*. Once a word has been borrowed from its source language into its recipient language, it gradually takes on the colour of that language, a process also referred to as *loan word accommodation*. The reason words are inclined to fall into the fountains of other languages is *language contact*, or contact between two or more languages, which can trigger change in those languages. Although less poetically formulated, this PhD thesis is concerned with the same phenomenon as introduced by Grimm: the borrowing and accommodation of loan words as a result of language contact. More specifically, we will assess whether the integration of loan words is constrained in different parts of speech, and if so, how those constraints on loan words manifest themselves. Additionally, we will investigate whether such constraints can *indirectly* impact the grammar of the recipient language.

Studies on language contact and borrowing are manifold. An example is the work by Wohlgemuth (2009; also see Wichmann & Wohlgemuth 2008), which heavily features in this thesis. Based on typological research, Wohlgemuth (2009) has described four *loan verb accommodation strategies*: direct insertion, indirect insertion, the light verb strategy, and paradigm insertion (as defined in Section 2.3.1.). The most common strategy across languages is *direct insertion*, where loan words can directly be implemented in their recipient language by adding the same inflections as to recipient-language words. Since direct insertion involves the addition of recipient-language inflections, and it is so frequent cross-linguistically, Wohlgemuth (2009: 291) has concluded that inflection does not hinder loan word accommodation. This refutes the ideas by, for instance, Harris & Campbell (1995:

135) and Sijs (2005: 56–57), who have argued that inflection forms the greatest barrier to loan word integration.

Wohlgemuth's (2009) ideas also prominently occur in the work by De Smet (2014), who has investigated the accommodation of English loan verbs in Present-day Dutch. He has found that the loan verbs can be integrated into the syntactic patterns and inflectional paradigms of Dutch by means of direct insertion (cf. Wohlgemuth 2009). In example (1), for instance, English loan verb *pushen* ('to push') is integrated in Dutch by directly adding Dutch inflections to loan verb stem *push*. Whereas the *-t-*inflection marks the past tense, the *-en-*inflection marks the plural, hence *push-t-en*.

(1)     *Ze* **pushten** *elkaar jarenlang naar eenzame hoogte*

         'They pushed each other to lonely heights for years.' (Twitter, 7 August 2021)

Inflectionally, *push-t-en* perfectly resembles native Dutch verbs used in the past plural, such as *fiets-t-en* ('cycled'), *stap-t-en* ('stepped'), and *werk-t-en* ('worked'). Despite the apparent ease of direct insertion as an accommodation strategy, De Smet (2014) has found that loan verbs cannot be integrated in all syntactic and inflectional categories with the same ease: there exist three tendencies in which loan verbs diverge from native verbs.

First, loan verbs are significantly more common in compound verb forms compared to native verbs, and this is the strongest tendency. Compound verbs in Dutch typically consist of an auxiliary verb, which is inflected, and a non-finite verb form (either an infinitive or a past participle). Example (2) illustrates how English loan verb *testen* ('to test') in Dutch is used as an infinitive in a compound verb. It is accompanied by auxiliary verb *kunnen* ('can'), inflected in the second person of the singular. Example (3) presents the use of English loan verb *focussen* ('to focus') in Dutch as a past participle in a compound verb. *Gefocust* is used with the inflected copular verb *is* ('is').

(2)     *en dan <u>kun</u> je je apparatuur* **testen** *uh of die uh of die nog juist is.*

         'And then you can test your equipment [to check], um, whether it, um, whether it is still all right.' (CGN)

(3)     *daar <u>is</u> 't helemaal op* **gefocust**.

         'It is entirely focused on that.' (CGN)

This tendency towards compound verbs suggests that loan verbs are significantly overrepresented in non-finite verb forms, and that they are underrepresented in finite verb forms compared to native verbs.

The other two tendencies found by De Smet (2014) relate to inflection: loan verbs are considerably more common in forms with fewer inflections compared to native verbs. In the present tense, for instance, loan verbs are more frequently used in the uninflected present singular, as in example (4), than in the *-t*-inflected present singular, as in example (5).

(4)    *'t enigste nadeel is soms uh als er zo organisaties zijn van uh theaterinstellingen of wat dan ook die rekenen daar dan op dat ik regelmatig mijn e-mail* **check**.

        'The only disadvantage is that sometimes, um, when there are like organisations of, um, theatre companies or, um, whatever, they count on me checking my emails regularly (lit. they count on it that I regularly check my e-mail).' (CGN)

(5)    *was wel vaag ja die spacecake [...] 't ging toch wel harder dan uh als je gewoon* **blow<u>t</u>** *of zo.*

        'It was quite blurry yeah that space cake [...] it went much faster than, um, if you had just smoked pot or something.' (CGN)

Based on the above three tendencies, De Smet (2014) has subsequently argued that the inflectional integration of loan words CAN form an obstacle to borrowing. This is a small yet crucial correction to Wohlgemuth's (2009) reasoning that inflection cannot limit verbal borrowing, as it is so commonly added to loan words cross-linguistically.

## 1.2. Aims and implications

In this research, we will expand the findings by De Smet (2014) from English loans in Dutch to French loans in Middle English (also called Medieval English), and from verbs to adjectives. Notice that the analyses for the Middle English data will address the later period, unless otherwise mentioned. The two overarching research questions are:

(i)    Are English loan words in Present-day Dutch and French loan words in Middle English biased towards specific morphosyntactic contexts in their recipient language? How do such biases manifest themselves, and are they persistent through time? Additionally, what factors are responsible for the occurrence of accommodation biases? (Part II)

(ii)    Do accommodation biases have long-term impact on the grammar of their recipient language? More specifically, has the great influx of French loans in English had lasting impact on some of the long-standing trends in the history of English? For instance, has the non-finite bias in French loan verbs promoted the overall use of non-finite forms in English? (Part III)

To answer those research questions, we will conduct corpus research on the distribution of loan words and native words[1] in specific grammatical and inflectional categories (cf. De Smet 2014). Data will be retrieved from linguistic corpora. We will manually mark the lexical tokens in the datasets, and make a two-way distinction between native words (i.e. Present-day Dutch and Middle English) and loan words (i.e. Present-day English and Anglo French). All tokens will additionally be annotated for a number of variables, such as grammatical category (for verbs), syntactic position (for adjectives), and inflectional endings.

In both contact settings, we will find constraints on loan word accommodation, which we will call *loan word accommodation biases*. The nature of these biases will suggest that loan words are overrepresented in categories where they carry less grammatical information, either because they carry no (or only few) inflections, or because the integration is achieved through a separate function word, namely the auxiliary or copula. This will allow us to confirm the correction to Wohlgemuth's (2009) argument as made by De Smet (2014). However, it should be noted that inflection is not the strongest barrier to accommodation, since the major biases are of a syntactic nature. Additionally, we will obtain a better understanding of the long-lasting persistence of accommodation biases by mapping the distributional properties of French-origin adjectives in Early and Late Middle English. We will also suggest which causes can be held accountable for the use of accommodation biases. Although we do not exclude the possibility of an interplay with other causes, we will hone in on one: the increased processing cost (as defined in Section 5.2.) when integrating loan words in recipient-language syntax. Accommodation biases will be proposed to serve as a facilitative strategy in diminishing that cost: biases mainly involve periphrastic structures, where auxiliaries and copulae carry most of the grammatical information. In this way, loan words carry less grammatical information, and they are less costly than in non-periphrastic structures. This reasoning will additionally allow us to digress on the long-standing debate on the dichotomy between borrowing and code-switching (or other-language insertions, as defined in Section 5.4.1.). More concretely, we will focus on lexical borrowings and single-word code-switches. Since accommodation biases reflect characteristics of both borrowing and code-switching, we will suggest that borrowings originate from single-word code-switches, and that the phenomena involved are tightly connected. Additionally, we will emphasise that the boundaries between borrowing and code-switching are unclear (cf. among others Thomason 2001: 133; Gardani 2008: 20; Matras 2009; Boas & Pierce 2011; Stammers

---

[1] Harley (2007) has shown that monolingual speakers may not treat all verbs equally. Some English verbs of Latin origin, for instance, still show systematic syntactic restrictions, as if they were still Latin verbs. However, in this thesis we artificially consider the bulk of native words a uniform group of words.

& Deuchar 2012: 630; Backus 2014: 103; Zenner & Geeraerts 2015: 248; Zenner 2017: 238; Zenner, Backus & Winter-Froemel 2019: 4).

After demonstrating the presence of loan word accommodation biases in two different contact settings, we will zoom in on the consequences of biases on their recipient-language grammar. The focus will be exclusively on verbs in the French-Middle English contact setting. Although most linguists agree that French has had a massive impact on the English lexicon, they have traditionally dismissed the possibility that French has also had an impact on Middle English syntax (e.g. Fischer 2013 and proponents). However, loan word accommodation biases associated with lexical *French verbs* being integrated in *English phrases* may have strengthened and accelerated already ongoing language-internal developments in English (cf. Hockett 1969: 414; Thomason 2001: 62). Constructions such as *do*-support, gerunds, light verb constructions, and infinitives are known to have developed in the course of the Middle English period, and they are considerably more common in Present-day English than in the continental West-Germanic languages (e.g. McWhorter 2002; Everaert 2008; Verhagen 2009; Berg 2014). The sharp increase of non-finite verb forms in Late Middle English and its aftermath coincided with the dramatic peak of French lexical items, and we will show that French influx may have played a role in the strong rise of non-finites. To this end, we will compare the distribution of loan verbs and native verbs in two candidate constructions: with *do*-support and in light verb constructions, two verbal periphrastic constructions which were on the rise in Late Middle English and Early Modern English. In both constructions, French-origin verbs will turn out be significantly overrepresented as compared to English-origin verbs. This does not mean that French influx alone can account for the sharp rise of *do*-support and light verb constructions: the literature and our findings show that a variety of internal and external motivators are at play in the emergence of both constructions. Instead, we will suggest that the accommodation of French loan verbs may have played its part in accelerating the developments. This implies that loan word accommodation biases can interact with ongoing developments and expansions in their recipient language, and that they can exert lasting influence on their recipient-language grammar.

By addressing these two research questions, this thesis assembles and advances the frameworks of contact linguistics, historical linguistics, and cognitive linguistics. To our knowledge, no one has recognised to date that loan word accommodation through direct insertion may come with constraints. This research will, therefore, contribute in conveying a more nuanced version of direct insertion (Wohlgemuth 2009), which is a differentiated phenomenon. The role of language contact in the rise of Middle English periphrastic constructions is largely unexplored terrain as well. The present research will be first in suggesting that French influx —

and particularly of verbs — may have *indirectly* impacted some of the long-standing trends in the history of English. Therefore, we will show that some changes involve no direct transfer, but arise as a result of transfer, namely through the morphosyntactic biases attached to the integration of lexical loans. Additionally, this research will refine our understanding of the history of the English language, and the controversial role which French has played in promoting some major trends in English. Those trends — triggered by contact with French — have caused the syntactic structure of English to diverge from the structure of the continental West-Germanic languages.

## 1.3. Structure of this thesis

This thesis will feature the following three parts.

Part I will present a state of the art on borrowing as a result of language contact, including relevant information on the accommodation and conventionalisation of loan words in their recipient language. Special attention will be paid to the four strategies of loan verb accommodation as described by Wohlgemuth (2009). Next, we will hone in on language change in the recipient language as a result of language contact, also called *contact-induced language change*, and on the internal and external factors which impact the type and extent of change. This will lead to a discussion of the two contact settings under investigation in this thesis: the English-Dutch and French-Middle English ones. For both contact settings we will provide a detailed description of the (historical) context as well as the intensity of contact. In a last step, we will present the general aims and hypotheses of this thesis.

In Part II we will develop a greater understanding of the phenomenon of loan word accommodation biases by addressing the first research question (cf. Section 1.2.). The existence, nature, and persistence of probabilistic biases will be discussed in Chapters 3. and 4., and their possible causes in Chapter 5. Whereas the main studies in Chapters 3. and 4. will focus on the later period of Middle English, Chapter 4. will additionally compare the distributional properties of French loan adjectives in Late and Early Middle English. It will become clear that biases are remarkably persistent throughout the Middle English period. After carefully documenting the tendencies of loan verbs and adjectives to be used in periphrastic and uninflected forms, Chapter 5. will demystify the potential causes of biases.

In Part III, then, we will deal with the potential long-term consequences of loan word accommodation biases on the grammar of their recipient language. To this end, we will address the second research question (cf. Section 1.2.), and we will rely on contact between French and Middle English to explain two syntactic developments in the Late Middle English period and its immediate aftermath. Chapter 6. will present the

findings for *do*-support, which came to a strong rise in Early Modern English. Chapter 7., next, will present a case study on light verb constructions, which came to a dramatic rise in the Late Middle English period. We will suggest that French influx may have considerably accelerated the rise of both constructions at the time of most intense contact.

The findings of this thesis will open various new avenues for future research, which will be discussed extensively in Chapter 8. in Part IV. Additionally, we will highlight some caveats relating to the data and methods as well as to the findings of this thesis. Last, we will expand on the general implications and scientific contribution of this PhD thesis.

# Part I

# Borrowing in contact

# Chapter 2

# Borrowing in contact

## 2.1. Introduction

This state of the art first comments on the process of borrowing as a result of language contact (Section 2.2.). After providing a definition of borrowing (Section 2.2.1.), we will introduce the concept of borrowability (Section 2.2.2.), and the hierarchies of borrowability (Section 2.2.2.1.) traditionally distinguished in the literature. Next, we will discuss possible reasons offered for borrowing (Section 2.2.3.), and three different types distinguished in the literature (Section 2.2.4.), namely lexical borrowing (Section 2.2.4.1.), syntactic borrowing (Section 2.2.4.2.), and morphological borrowing (Section 2.2.4.3.), with a focus on lexical borrowing. Since loan words have to be accommodated to their recipient-language structure, Section 2.3. will describe the process of loan word accommodation. This will allow us to expand on the four cross-linguistic loan verb accommodation strategies (Section 2.3.1.) as identified by Wichmann & Wohlgemuth (2008) and Wohlgemuth (2009): direct insertion (Section 2.3.1.1.), indirect insertion (Section 2.3.1.2.), the light verb strategy (Section 2.3.1.3.), and paradigm insertion (Section 2.3.1.4.). Loan words also become conventionalised (as defined by Schmid 2015) and, therefore, shared by the speech community, a process which will be discussed in Section 2.3.2. The focus of Section 2.4. will be on language contact settings. In a first step, we will introduce the phenomenon of contact-induced language change (Section 2.4.1.), and we will not only cover direct transfer, but also indirect transfer, which is widely understudied (Section 2.4.1.1.). Factors affecting the outcome of contact settings will be elaborated on as well (Section 2.4.2.). In a second step, we will discuss the two contact settings under investigation in this thesis, namely the French-Middle English contact setting (Section 2.4.3.) as well as the English-Dutch contact setting (Section 2.4.4.). The focus will be on Middle English, which has left behind a vast body of texts, and which is the focus of this thesis. After a detailed description of the (historical) context and the intensity of contact of both settings, we will identify direct insertion as the loan word accommodation strategy which is most commonly used in both settings. We will, then, draw a comparison between the two contact situations (Section 2.4.5.) based on

the factors highlighted in the literature. We will conclude this state of the art by describing the two main aims and hypotheses of this thesis (Section 2.5.).

## 2.2. Borrowing as a result of language contact

The phenomenon of *language contact* can be defined as "the use of more than one language in the same place at the same time" (Thomason 2001: 1), but "some communication between speakers of different languages is necessary" (Thomason 2001: 1). Language contact is particularly common, since it "occurs in nearly every country of the world and is manifest in the documented history of most languages" (Gardani 2008: 3). Apart from being common, it is also a particularly old phenomenon, since language contact has existed from the moment humans have started to speak more than one language (Thomason 2001: 6). What is more, it has never been proven "that any languages have developed in total isolation from other languages" (Thomason 2001: 8). Following Weinreich's (1953: 1) pioneering work, language users are the locus of this contact, which is also mirrored in Bowern's (2013) "[c]ontact is what speakers do, not what languages do". Although the study of language contact is important, Campbell (1998) has warned to be careful before attributing every possible change to foreign influence.

Following Whitney (1881: 10), "wherever two tongues come in contact, each is liable to borrow something from the other". In other words, language contact can result in borrowing, which, in this thesis is defined as the phenomenon in which an item has been copied from a language X into a language Y at some point in time. Items which are borrowed are gradually integrated morphosyntactically in language Y, and become highly established and conventionalised over time. Language contact research is characterised by a persistent debate on the relationship and differences between lexical *borrowing* and *code-switching*. The phenomenon of code-switching occurs when insertions from a language X are used in a language Y within the same sentence. Code-switches are typically uttered by multilingual speakers, and in this thesis the focus is on switches between two languages only[2]. In both lexical borrowing and code-switching, other-language material is inserted in a recipient (for borrowing) or matrix (for insertional code-switching) language. However, researchers do not agree on the exact definitions of the phenomena, nor on their mutual relationship. For instance, in which cases is inserted other-language material a code-switch, and in which cases is it a borrowing? This question especially applies to single-word insertions. In this thesis we will distinguish between code-switching and borrowing by means of the morphosyntactic integration and conventionalisation or usage frequency of the inserted material: whereas code-switching is characterised by

---

[2] We acknowledge that more than two languages can be involved in language contact, but this does not fall under the scope of this thesis.

non-integration and lower levels of conventionalisation, borrowing is characterised by integration and higher levels of conventionalisation. Another oft-repeated topic of debate is whether one practice originates from the other, or whether they are entirely unrelated. That is, while some researchers consider code-switching and borrowing virtually different names for the same phenomenon (e.g. Torres Cacoullos & Aaron 2003; Schendl 2004; Onysko 2007; Poplack & Dion 2012), other researchers do not see any connection between the two phenomena at all (mainly Poplack and colleagues, e.g. Poplack, Sankoff & Miller 1988; Poplack, Wheeler & Westwood 1989; Poplack & Meechan 1999; Poplack 2017: 201; Poplack et al. 2020). Yet other researchers suggest that borrowing and code-switching are naturally linked — hence not identical — and that they are ends of the same continuum (e.g. Myers-Scotton 1993b; Treffers-Daller 2005; Matras 2009; Backus 2015). A more elaborate definition of code-switching as well as an in-depth discussion of the long-standing debate on the dichotomy between code-switching and borrowing will be presented in Section 5.4. However, this state of the art focuses solely on borrowing[3].

## 2.2.1. Definition and terminology

> It is common for one language to take words from another language and make them part of its own vocabulary: these are called loanwords[4] and the process is called linguistic *borrowing*. (Campbell 1998: 57)

The phenomenon of borrowing (also called mixture, transmission, or replication) typically involves material which has been copied from a language into another language at some point in the past. The language from which the material originates is called the *source language* (also called donor language, lending language, or model language), and the language which adopts it is called the *recipient language* (also called target language, host language, borrowing language, or replica language). Note, however, that this traditional terminology as used by Weinreich (1953) and proponents has been criticised by Johanson (2002, 2008; also see Hock & Joseph 2019: 223) in his code-copying framework. Johanson's (2002: 288) argument is that

> [t]he term "borrowing" is already based on a deceptive metaphor. Nothing is borrowed in language contact: the "donor language" is not deprived of anything; and — more importantly — the "recipient language" does not take over anything identical with anything in the "donor language".

---

[3] An exception is the discussion of historical code-switching in Medieval England (Section 2.4.3.1.3.).

[4] This thesis does not expand on other types of borrowed elements, such as loanblends (Haugen 1950: 214), loanshifts (Haugen 1950: 220), and loan formation (Onysko 2007).

For example, the word *restaurant* in English (and many other languages) comes from French, and now belongs to both French and English. Also, transferred words are never identical to their original source-language words, as they may differ in spelling, pronunciation, or grammatical properties (Johanson 2008: 62–63). Therefore, Johanson (2008: 62–63) has proposed to use the term 'copying' instead of 'borrowing'. We agree with Johanson's (2002: 288) reasoning that borrowings are not literally 'borrowed', since they continue to exist in their source-language lexicon; however, this thesis will adopt the traditional terminology, as it is used by most authors cited in this state of the art. On a side note, following Haugen (1950: 211) the term 'language mixture' will be avoided throughout this thesis, as it implies language purism and may, therefore, be interpreted pejoratively. Also note that the term 'loan word' (or simply 'loan') in this thesis generally refers to the word which is being borrowed, and 'borrowing' to the process in which loan words are being borrowed (cf. definition provided by Campbell 1998: 57). However, whenever we explicitly contrast loan words to code-switches, we will use not use the term 'loan words', but 'borrowings' (cf. Chapter 5).

In the literature, borrowing is considered a natural phenomenon (Janssens & Marynissen 2008: 172). According to Whitney (1881: 10), for instance, "wherever two tongues come in contact, each is liable to borrow something from the other". Borrowing is also a highly common phenomenon: "[i]n virtually every bilingual situation empirically studied, borrowed items make up the overwhelming majority of other-language material" (Poplack 2018: 1). At the same time, borrowing is a particularly complex phenomenon, and Rothwell (1980: 118) has stressed that to explain why an element is borrowed "at a particular moment and in a particular place" a set of factors has to be taken into account. The definition and features of borrowing are further discussed in Section 5.4.2. Also note that this thesis adopts a user perspective of loan words (cf. Matras 2013: 11), or

> an approach to contact linguistics that regards languages less as static systems, and more as dynamic repertoires, and speakers not just as followers of social norms, but as creative contributors to the shape of linguistic structures and routines.

## 2.2.2. Borrowability
Elements of the language vary in terms of *borrowability*: following Wohlgemuth's (2009: 54) definition,

> [some] parts of the lexicon and grammar appear to be generally more resistant to borrowing than others. In other words, they vary with respect to the frequency they are being borrowed and the ease with which they are accommodated.

So far, the focus of research on borrowability has mainly been on single words or parts of speech, and not so much on multiple words or chunks of text (Doğruöz &

Zenner 2013). This thesis will follow this more traditional approach of focusing on parts of speech.

### 2.2.2.1. Hierarchies of borrowability

Thomason (2001: 63) asserts that anything is borrowable. Borrowability tends to be measured in terms of *hierarchies of borrowability* (also called borrowing hierarchies or borrowing scales), or hierarchies which indicate which elements or parts of speech are *more borrowable* than others. For example, researchers generally agree that lexical items — and specifically content words — are more borrowable than syntactic items and function words (e.g. Field 2002; Winford 2003; Matras 2009). However, when using the phrase 'X is more borrowable than Y', one should bear in mind that borrowability hierarchies have at least four possible interpretations (i.e. a temporal, implicational, quantitative, and probabilistic interpretation), and that one should be specific about which interpretation is aimed at (Haspelmath 2008: 48). Eisen (2019: 51) has explained the meaning differences between the possible interpretations — and the subsequent need for clarity — by means of a concrete example. A temporal interpretation of *Nouns are more borrowable than adjectives*, first, is that nouns are borrowed before adjectives. Implicationally, the same phrase would be interpreted as *If the noun is borrowed, then the adjective will be borrowed as well*. Third, a quantitative interpretation of *Nouns are more borrowable than adjectives* would be that there is a higher number of borrowed nouns than adjectives in the language. Probabilistically, the interpretation of *more borrowable* is that nouns are generally more likely to be borrowed than adjectives. In the continuation of this thesis, each instance of *X is more borrowable than Y* should be interpreted quantitatively-probabilistically, unless explicitly described otherwise.

In his seminal work on "mixture" in language, Whitney (1881) has proposed the following hierarchy of borrowability:

> nouns > adjectives > verbs > adverbs > prepositions and conjunctions > pronouns > derivative prefixes and suffixes > inflectional prefixes and suffixes > sounds[5]

Following this hierarchy, the most borrowable part of speech is nouns, followed by adjectives and verbs. The same tendency for nouns to be the most borrowable part of speech has been found in code-switching (Poplack 1980). Although verbs are generally less borrowable than nouns, they are more borrowable than prototypical closed-class items such as affixes and conjunctions (Muysken 1981; Matras 2007), but in this thesis we solely focus on the borrowability of nouns, adjectives, and verbs. Interestingly, loan verbs often enter their recipient languages as items other than

---

[5] It is important to bear in mind that the parts of speech mentioned in Whitney's (1881) hierarchy of borrowability may not apply to all languages, and that parts of speech in different languages may be incompatible. For example, not all languages have clearly delineated categories such as adjectives and nouns.

verbs, for instance as nominal forms, which are then gradually converted into verbal forms (Whitney 1881; also see Moravcsik 1975a). Other researchers, such as Hock & Joseph (2019: 227), have noticed that,

> if the need for borrowing does arise, many languages instead borrow a nominal form of the verb and employ a native all-purpose verb such as *do* or *make* as a means of turning that form into the equivalent of a verb.

This is reminiscent of Wohlgemuth's (2009) *light verb strategy*, where an inflected light verb is combined with a deverbal noun, and which will be introduced in Section 2.3.1.3. The fact that verbs are less borrowable than nouns is a view generally shared by linguists (e.g. Muysken 1981; Poplack, Sankoff & Miller 1988; Matras 2007, 2009; Winford 2010; Hock & Joseph 2019), but Wohlgemuth (2009: 291) has not found evidence for this claim. Admittedly, loan nouns are partially more common than loan verbs because nouns are generally more common than verbs (Matras 2009: 157; Wohlgemuth 2009: 250), but even with this frequency difference taken into account nouns are generally more borrowable than verbs (Hout & Muysken 1994; Matras 2007). A handful of reasons which have been provided to explain this difference in borrowability is that verbs carry more inflections than nouns (Harris & Campbell 1995), nominal inflections are more frequently borrowed than verbal inflections (Seifart 2019), verbs are morphologically more complex than nouns (Meillet 1921: 175; Matras 2007, 2009; Winford 2010: 178), and verbs are more functional than nouns (Haugen 1950). Additionally, verbs are less open class than nouns, and closed-class items are generally less likely to be borrowed than open-class items (Muysken 1981). The borrowability of nouns and verbs in the literature is typically discussed more often than the borrowability of other parts of speech, such as adjectives (cf. Stammers & Deuchar 2012).

The traditional hierarchies of borrowability have also been criticised, for instance by Romaine (1989: 63–66), who has found that — in the Panjabi/English community in Britain — verbs are borrowed more frequently than nouns. Therefore, the traditional hierarchy cannot account for all languages in the world. Poplack, Sankoff & Miller (1988) have argued that hierarchies of borrowability reflect the division of syntactic categories in the recipient language rather than the actual borrowing rates of specific parts of speech (cf. supra).

## 2.2.3. Reasons for borrowing

Borrowing is a common phenomenon; however, it only happens when "one […] see[s] a role for the alien item in one's native language" (Strang 1970: 28). The reasons for borrowing offered in the literature are manifold. A frequently cited reason is that the source language may be culturally dominant. By using words from the source language, speakers may be associated with speakers of that dominant language and,

therefore, gain prestige (Field 2002: 4; Matras 2009). An example is the use of "Latin phrases in English" (Weinreich 1953: 60), Latin being considered the more prestigious language. Indeed, Johanson (2002) has observed that more prestigious (also called superstrate) languages tend to influence less prestigious (also called substrate) languages. Another example can be found in isiNdebele, a South African language which has borrowed extensively from English due to the prestige ascribed to English (Mahlangu 2016). In the contact settings under investigation in this thesis prestige plays a crucial role as well: French in Medieval England was a high-prestige language (cf. Section 2.4.3.), and English in the Low Countries is associated with social prestige (cf. Section 2.4.4.). Although the above examples all involve more prestigious languages influencing less prestigious languages (*superstrate influence*), this does not imply that less prestigious languages cannot influence more prestigious languages as well (*substrate influence*), as shown by Winford (2013). That said, substrate influence is cross-linguistically less common than superstrate influence.

A second reason commonly used to account for borrowing is the need to fill lexical gaps (Rothwell 1980: 118; Field 2002: 4). Weinreich (1953: 56), for instance, has stressed "[t]he need to designate new things, persons, places, and concepts". In the absence of a native word to refer to an item or concept, a word is borrowed from another language. Such loans have also been called *necessary loans* (also called core loans) in the literature. In Dutch, for example, many English loan words (e.g. *compact disk*, *computer*) are IT-related since many IT developments have emerged in Anglo-Saxon countries. Necessary loans are opposed to *luxury loans*, which have an existing synonym in their recipient language, but which have more prestige than those existing recipient-language equivalents. Weinreich (1953: 57) has additionally stressed the economic advantage of borrowing, since "[u]sing ready-made designations is more economical than describing things afresh". A last explanation is that sometimes words are borrowed to compensate for the low frequency of the native word, which — due to its low frequency — is "more subject to oblivion and replacement" (Weinreich 1953: 57)[6]. Therefore, the native word is replaced by a loan word. Although this list is by no means exhaustive, it provides an overview of frequently mentioned reasons for borrowing offered in the literature.

## 2.2.4. Types of borrowing

According to Culpeper (2005: 35), "[b]orrowing occurs at all levels", and Campbell (1998) has stated that it is not restricted to lexical items, but instead expands to sounds, phonological rules, grammatical morphemes, syntactic patterns, semantic associations, and discourse strategies. Thomason (2001: 11) as well has confirmed that "all aspects of language structure are subject to transfer from one language to another,

---

[6] Weinreich (1953: 57) does not give examples of this phenomenon.

given the right mix of social and linguistic circumstances". This thesis presents *lexical* (Section 2.2.4.1.) and *syntactic borrowing* (Section 2.2.4.2.), which have been traditionally distinguished in the literature (e.g. Thomason & Kaufman 1991: 77; Matras 2009). Additionally, we discuss *morphological borrowing* (Section 2.2.4.3.), a concept which has been coined by Gardani (2008: 95–96). In those three types of borrowing, transfer can either be positive (i.e. when a feature is added due to contact) or negative (i.e. when a feature is lost due to contact) (Gardani 2008: 22).

### 2.2.4.1. Lexical borrowing

Lexical borrowing (also called lexical transfer[7]) can be defined as "the process whereby words from a lending language [i.e. source language] become entrenched as conventional words in the receiving [i.e. recipient] lexicon" (Backus & Dorleijn 2009: 77). Lexical borrowing typically involves single-word insertions. Coetsem (1988: 98) has distinguished two types of lexical borrowing, namely *lexical addition* and *lexical replacement*. Although he has pointed out that those terms involve more than just addition and replacement (Coetsem 1988: 98), we do not expand on this subdivision.

Lexical borrowing is traditionally considered to "take[…] place most frequently and effortlessly [compared to syntactic borrowing] because lexical items can be easily detached from the donor language, even when a speaker is unfamiliar with this language" (Fischer 2007: 27; also see Field 2002). Weinreich (1953: 56) has indeed characterised vocabulary as the "domain of borrowing *par excellence*", since "lexical borrowing is less restricted to the bilingual portion of a language community than phonic or grammatical interference". Moreover, according to Seifart (2019: 16), almost all languages in the world have lexical borrowings. The ease of lexical borrowing becomes clear from the fact that it is attested in casual language contact settings where contact is not necessarily intense (Thomason & Kaufman 1991: 74; Bowern 2013: 419). English, for instance, has borrowed a number of words from Japanese, such as *haiku*, *karaoke*, and *zen* (cf. OED), yet Japanese is not an official language in any of the English-speaking countries. However, according to Thomason & Kaufman's (1991: 74; also see Thomason 2001: 70–71) *borrowing scale*, when contact is casual only content words will be borrowed, and in particular non-basic vocabulary. Function words, such as conjunctions and adverbial particles, will be borrowed only under slightly more intense contact. When contact is even more intense, other function words, such as adpositions, affixes, pronouns and numerals, may be borrowed as well.

In this thesis the focus is on borrowings of the lexical type, which are accommodated to their recipient-language grammar (Section 2.3.). Also note that there is a tendency

---

[7] In this thesis 'borrowing' and 'transfer' will be used interchangeably.

in historical research to mainly investigate the lexical borrowing of nouns (Timofeeva & Ingham 2018: 202); the research presented in this thesis is innovative in that the focus is on the lexical borrowing of verbs and adjectives.

### 2.2.4.2. Syntactic borrowing

Syntactic borrowing (also called grammatical and structural borrowing, or syntactic transfer) can be defined as

> the process whereby the use of a structure originally from the donor [i.e. source] language becomes entrenched as a conventional part of the grammatical structure of the receiving [i.e. recipient] language. (Backus & Dorleijn 2009: 78)

Wohlgemuth (2009: 224) defines syntactic borrowing as "those cases where the transferred grammatical morpheme retains its function — or one of its functions — in the recipient language". Examples of syntactic borrowings from a wide range of languages have been provided by Matras & Sakel (2007), and examples specifically for the French-Middle English contact setting will be discussed in Section 2.4.3.2. In cases of particularly intense language contact, the outcome may be a pidgin or creole language (Romaine 1989: 66), as defined in Thomason (2001: 159–162).

The frequency of syntactic borrowing is much more contested than that of lexical borrowing. Different from lexical borrowing, which can occur in casual language contact settings, Bowern (2013: 420) has found that syntactic borrowing occurs in contact settings characterised by intense contact stretching over a long period. This also becomes clear from Thomason & Kaufman's (1991: 74–75) borrowing scale, where syntactic borrowing is suggested to occur in contact situations ranging from "[s]lightly more intense contact" to "very strong cultural pressure" (also see Muysken 1981; Coetsem 1988: 25; Thomason 2001: 70–71; Thomason 2016: 43). *Cultural pressure* in that sense is defined as "any combination of social factors that promotes borrowing", examples being "prestige or economic forces that make bilingualism necessary" (Thomason 2016: 76). Additionally, words are said to be borrowed first, and grammar later (Thomason 2001: 64). Therefore, it has traditionally been proposed that syntactic borrowing is rare (e.g. Meillet 1921: 87; Weinreich 1953). Hockett (1969: 415; also see Meillet 1921: 87) has agreed that *direct* syntactic borrowing is rare, but possible. However, he has also suggested that

> grammatical change can be brought about indirectly by borrowing–via sets of related loanwords. There is no doubt that grammatical change can result from borrowing *from another language* in any other way. (Hockett 1969: 414)

The option of *indirect* syntactic transfer will be extensively discussed in Section 2.4.1.1.

Going against tradition, some researchers have observed that syntactic borrowing is

> much more frequent and important than some scholars have thought in the past, though others have gone to the other extreme of assuming that everything not otherwise readily explained in a language's grammar is due to borrowing. (Campbell 1998: 230)

One of the reasons as to why there is "much less agreement on the extent of syntactic borrowing" (Fischer 2007: 28) may be that borrowing at the syntactic level is harder to recognise than borrowing at the lexical level (Thomason 2001: 91; Fischer 2007: 26). The existence and extent of syntactic borrowing is controversial in, for example, research on the impact of French on Middle English syntax. Traditionally, linguists (mainly Fischer's followers) have claimed that French has not (or only to a limited extent) exerted syntactic influence on the English language (e.g. Lass 1987: 54–61; Thomason & Kaufman 1991: 307–310; Townend 2006: 70; Fischer 2013). However, more recently a group of linguists (mainly Rothwell's and Ingham's followers) has found that the rate of structural borrowings has been largely underestimated in the literature, and that the traditional assertion should be re-evaluated (e.g. Rothwell 1968, 1976, 1998; Trotter 2003; Ingham 2006, 2009a, 2012, 2020; Haeberli 2010; Trips 2014). This topic will be discussed more prominently in Section 2.4.3.2.

### 2.2.4.3. Morphological borrowing

Morphological borrowing (also called inflectional borrowing or morphological transfer) is a fairly recent term in contact linguistics, its study being much less advanced than that of lexical borrowing (cf. Gardani 2008). Gardani (2008: 95–96) has been the first linguist to provide a definition for morphological borrowing:

> the shift of inflectional morphemes from a source language B into a receiving [i.e. recipient] language A on the part of a bilingual speaker of A, whereby they are added *only* to native words of the receiving language and have maintained the (at least partially) identical meaning (and function) they carried out in the source language, once they entered the receiving language.

Ever since Gardani (2008), the field has started to receive more attention (cf. Gardani 2018). An example of a morphological borrowing can be found in Malinche Nahuatl (Gardani 2018: 9–10), an American-Indian language spoken in Central and Western Mexico (Encyclopædia Britannica n.d.). Malinche Nahuatl has adopted the use of Spanish augmentative formative *-ote* (Gardani 2018: 9–10), as in *guap-**ote***, where *guapo* means 'attractive' and *-ote* 'very', hence 'very attractive' (Gardani 2018: 9–10). In the Malinche Nahuatl language, Spanish-origin *-ote* can be found in, for instance, *huēy-**ote***, where *huēy* means 'big', and *huēy-**ote*** 'very big'. Another example is the productive use of French-origin adjectival suffix *-able* (from

Latin -*abilis*) in Middle English, as in *bilevable* ('believable') and *findable* ('findable') (Lewis 1952–2001).

Following Thomason (2016: 43), transfer of inflectional material is rare, "even in intense contact situations". That is because inflectional morphology is considered "one of the least borrowable parts of the language's structure" (Comrie 2008: 15), which reflects the traditional findings represented in hierarchies of borrowability (cf. Whitney 1881; Haugen 1950; Moravcsik 1975b; Matras 2007). However, according to Gardani (2018) researchers do not have an idea of how frequent morphological borrowing actually is.

Due to its scope, this thesis will not expand on the process of morphological borrowing any further.

## 2.3. Loan word accommodation

According to Weinreich (1953: 44),

> [a] word which has been transferred from one language into another is itself subject to the interference of the grammatical, as well as the phonic, system of the recipient language, especially at the hands of its unilingual speakers.

In short, loan words — after having been borrowed from their source language — "undergo various types of *post hoc* adaptation" (Andersen 2017: 127) to accommodate to their recipient language, and this is specifically true in the initial stage. As a result of *loan word accommodation* (or loan word adaptation, assimilation, or integration), a loan word "becomes an integral part of [the recipient] language, undistinguished, except to reflective and learned study, from the native material" (Whitney 1881: 16). The process of loans being integrated more strongly into the linguistic structure of their recipient language has also been called *nativisation*[8] by Hock & Joseph (2019: 229). When a construction or word is used productively (i.e. used more than once), it is called an *accommodation pattern*, or "a construction to accommodate a loan verb that is or was applied by the recipient language productively (i.e. in more than one occasional case)" (Wohlgemuth 2009: 68). Weinreich (1953: 44, 46) has pointed out that some loan words are not integrated at all, although accommodation is more frequent than non-accommodation.

The process of loan word accommodation has been described by Saugera (2012: 227) as "a gradient phenomenon, thus involving various phases through various linguistic processes and forms". Andersen (2017) has proposed that loan words can adopt

---

[8] In his Dynamic Model, Schneider (2007) has illustrated how English as a foreign language (in areas to which it was transported because of colonialism) sheds it foreignness in order to become an indigenous language. 'Nativisation' is the third out of five phases in this process. However, this thesis does not expand further on the Dynamic Model.

orthographic, phonological, morphological and semantic properties, and Zenner, Heylen & Van de Velde (2018) have additionally discussed pragmatic properties. Although the phonological integration of a loan word can reveal how integrated it is in its recipient language, this falls beyond the scope of this thesis. Instead, we fully focus on *morphosyntactic integration*, involving both syntactic integration (Weinreich 1953), or integration in the recipient-language syntactic system, and morphological integration, or integration in the recipient-language morphological system. Morphosyntactic integration includes, for example, the integration of loan words in the inflectional system and usage categories[9] of their recipient language. It may be argued that the different types of integration are — at least to a certain extent — in interaction. However, since Poplack et al. (2020) have suggested that morphosyntactic and phonetic integration are independent of one another, both can be safely looked at separately.

The exact steps in the loan word accommodation process are dependent on the part of speech. However, loan words of most parts of speech have in common that, when they enter their recipient language, they are generally assigned to a specific syntactic category (Poplack, Sankoff & Miller 1988). In a next step, they accommodate to the syntactic structures of their recipient language (e.g. Haugen 1950; Poplack, Sankoff & Miller 1988: 52; Muysken 2000: 184; Wichmann & Wohlgemuth 2008). The complexity of this phenomenon is now illustrated for the case of loan verbs entering into their recipient language. Loan verbs optionally undergo the following five processes[10], which have been literally reproduced from Wohlgemuth (2009: 56):

– assigning a loan verb to the word class 'verb'

– assigning it to an inflectional class

– assigning to it a classifying verb or an inflecting verb (in complex predicates)

– assigning valency to it

– attaching inflectional morphology to it

Wohlgemuth (2009) has added that other processes can occur in addition to the five attested here. An example which comes to mind is the "phonological shape of the borrowed item [which] may be made to conform with recipient-language patterns" (Poplack, Sankoff & Miller 1988: 62), but as mentioned above, phonological accommodation falls beyond the scope of this thesis.

---

[9] 'Usage categories' are in this thesis used to denote grammatical and inflectional categories.
[10] Wohlgemuth (2009: 56) does not specify whether loan words undergo these processes all at once, or whether they proceed through (diachronic) phases.

The next section (2.3.1.) zooms in on four accommodation strategies of loan verbs as identified by Wichmann & Wohlgemuth (2008) and Wohlgemuth (2009). Those strategies — and direct insertion in particular — constitute a fundamental base for the hypotheses formulated in this thesis. Section 2.3.2., next, will present how loan words become conventionalised in their recipient language after their integration, and how conventionalisation can be measured.

## 2.3.1. Loan verb accommodation strategies

In their typological research on loan word accommodation, Wichmann & Wohlgemuth (2008) and Wohlgemuth (2009) have identified four main loan word accommodation strategies which apply specifically to loan verbs: direct insertion (Section 2.3.1.1.), indirect insertion (Section 2.3.1.2.), the light verb strategy (Section 2.3.1.3.) and paradigm insertion (Section 2.3.1.4.). All four strategies — with the exception of the latter — are cross-linguistically used to accommodate loan verbs to their recipient language. The strategies differ in, among others, the type of transfer and the integrational effort needed. The observations were based on a Loan Verb Database (LVDB), compiled by Wohlgemuth (2009), containing data from 429 languages (Wohlgemuth 2009: 25). All strategies are discussed in detail in the subsections below and will be referred to throughout this thesis.

### 2.3.1.1. Direct insertion

The first strategy discussed here, *direct insertion* (Wohlgemuth 2009: Ch. 6.), is characterised as follows:

> the borrowed verb […] is immediately available for the grammar of the recipient language without any morphological or syntactic adaptation whatsoever being necessary to render the replica equivalent to a native verb (or verb stem). (Wohlgemuth 2009: 87)

In this strategy, native inflections can thus be added *directly* onto the loan verb stem. Based on Wohlgemuth's (2009) database, this accommodation strategy is used in 52.5% of the cases and is, therefore, the most frequent strategy across languages. Direct insertion mainly occurs in SVO languages, such as in most languages of Germanic and Romance descent, but it has also been attested in various non-Indo-European languages. That this strategy is so common is not surprising, since direct insertion requires little integrational effort[11] (Wohlgemuth 2009: 135). Instances of direct insertion are printed below: an English loan verb is integrated in

---

[11] Integrational effort is defined as "expenditure of any morphological, morphophonological, or morphosyntactic operation that is necessary to *adapt* a borrowed lexical item into the system of the recipient language" (Wohlgemuth 2009: 135).

German (6), and a Spanish loan verb in Aymara, an American language native to Bolivia, Chile, and Peru (7).

(6)     German

*download-en*
download- INF
'to download'

< English *to download* (Wohlgemuth 2009: 88, example based on his own data)

(7)     Aymara

*wiyaja-ña*
travel-INF
'to travel'

< Spanish *viajar* 'to travel' (Wohlgemuth 2009: 89, example based on Hardman, Vásquez & Yapita 1988: 55)

In both examples, recipient-language infinitival markers (*-en* and *-ña* respectively) are added directly onto the loan verb stems (*download-* and *viaja-* respectively). However, in the case of Aymara this involves a slight phonological modification[12] (i.e. initial *v* being converted into *w* and the addition of intervocalic *-y-*) in order to meet Aymara phonotactic requirements (Wohlgemuth 2009: 89).

Surprisingly, direct insertion is also attested in less compatible language pairs, for instance in cases where one language has richer verbal morphology than the other (Wohlgemuth 2009: 89). An example of this phenomenon is presented in (8), where a loan verb from Portuguese, a synthetic language, has accommodated into West Greenlandic, a polysynthetic language.

(8)     West Greenlandic

*paliaar-poq*
dance-3SG
'(s)he takes part in singing and dancing'

< Portuguese *bailar* 'to dance' (Wohlgemuth 2009: 89, example based on Voort 1995: 139)

Portuguese *bailar* has undergone slight phonological modification (*paliaar*), and receives the West-Greenlandic inflection *poq* in the third person of the singular.

---

[12] It is highly conceivable that English *download* in German *downloaden* undergoes phonological modification as well, such as the pronunciation of the vowels, but this cannot be verified by looking at the orthography.

Wohlgemuth (2009) has even encountered some attestations where direct insertion is applied to word classes other than verbs, such as nouns. For instance, Tuareg, a Berber language, has a number of "action-word nouns with more 'verby' semantics" (Wohlgemuth 2009: 90). Although they formally resemble nouns, they can easily be inserted into the verbal paradigm because of their meaning[13]. In example (9), action-word noun *tusrak* 'sneezing' is accommodated into Tasawaq, a Northern Songhay language, as a full-fledged verb.

(9)     Tasawaq

    *ghá b-tásrìg*
    1S IPFV[14]-sneeze
    'I am sneezing'

    < Tuareg *tusrak* 'sneezing' (Wohlgemuth 2009, example based on Wichmann 2004 and Maarten Kossmann p.c.)

The example shows that *tusrak* (*tásrìg* in Tasawaq) receives Tasawaq verbal inflections.

More examples of direct insertion are presented further on for the specific language contact settings discussed in this thesis (Sections 2.4.3.3. and 2.4.4.3.).

### 2.3.1.2. Indirect insertion

The second accommodation strategy is *indirect insertion* (Wohlgemuth 2009: Ch. 7.), where loan verbs are adapted morphosyntactically before accommodating to their recipient language:

> [Loan verbs] are inserted *in*directly, because this loan verb accommodation technique involves adaptation by overt (verbalizing) affixation of some kind. Once that affix is added, however, the borrowed word is a fully functional verb in the recipient language and normal inflectional patterns may be applied to it. (Wohlgemuth 2009: 94)

Since loan verbs in this strategy are not immediately available as fully functional verbs, the integrational effort is said to be higher than in direct insertion. Indirect insertion occurs in 20.6% of cases of verbal accommodation cross-linguistically, making it the third most frequently used accommodation strategy.

In order to indirectly insert loan verbs, many languages use specialised verbalizers, or "affixes whose sole or main purpose is verbalizing derivation" (Wohlgemuth 2009:

---

[13] Although this is reminiscent of the light verb strategy (cf. Section 2.3.1.3.), where deverbal nouns are integrated in the verbal paradigm, such cases of direct insertion differ fundamentally from the light verb strategy. Whereas the 'verby' noun in example (9) can be inflected, the deverbal noun in the light verb strategy cannot itself be inflected. Instead, inflections are added to the light verb.

[14] IPFV stands for 'imperfective' (cf. Wohlgemuth 2009: xxvii).

95). Hungarian, for instance, uses *-ál* and *-ol* in order to verbalize (both native and borrowed) nouns. This feature of indirect insertion is illustrated in example (10), where German loan verb *leisten* is accommodated to the Hungarian language by adding *-ol*.

(10)   Hungarian

*leiszt-ol*
accomplish-VBLZ
'to accomplish'

< German *leisten* 'to work hard, accomplish' (Wohlgemuth 2009, example based on Moravcsik 1975a: 5–7)

Only after adding the verbalizer, the loan verb becomes fully functional in its recipient language and can be used to its maximum potential: for instance, it can be used in all forms and persons, and it can be inflected like recipient-language verbs. This is shown in example (11), where English loan verb *to check* only becomes functional in Modern Greek after it has received Greek verbalizer *-ar*.

(11)   Modern Greek

*tsek-ar-i*
check-VBLZ-3SG
'(it) checks'

< English *to check* (Wohlgemuth 2009: 96, example based on his own data)

After receiving verbalizer *-ar*, the loan verb can receive a Modern Greek *-i*-inflection to denote the third person of the singular.

Other languages use factitives or causatives to operationalise verbs, including loan verbs. Example (12) illustrates the use of a factitive (*-in*) attached to an English loan verb (*download*) in Jakarta Indonesian.

(12)   Jakarta Indonesian

*download-in*
download-FACT
'to download'

< English *to download* (Wohlgemuth 2009: 97, example based on Tessa Yuditha p.c.)

A similar phenomenon, this time with the causative (*ī-*), is demonstrated for English *to miss* in Ma'di, a Central Sudanic language (13).

(13)    Ma'di

 *ī-mīsì*
 CAUS-miss
 'to miss (someone)'

 < English *to miss* (Wohlgemuth 2009: 97, example based on Blackings & Fabb 2003: 69)

In yet other cases, indirect insertion in loan verbs is realised by means of a distinct loan verb marker (LVM), "a distinct affix whose sole function is to accommodate borrowed verbs" (Wohlgemuth 2009: 98), as in example (14).

(14)    Sinte/Burgenland Romani

 *roas-in-av*
 travel-LVM-INF
 'to travel'

 < Bavarian German *roasn* 'to travel' (Wohlgemuth 2009: 98, example based on Bakker 1997: 6)

In (14), the Bavarian German verb *roasn* in Burgenland Romani receives loan verb marker *-in-* in addition to infinitival *-av-*marker.

In a limited number of cases, loan verb accommodation in indirect insertion is realised by yet other means (cf. Wohlgemuth 2009: 100).

### 2.3.1.3. Light verb strategy

The third strategy, the *light verb strategy* (Wohlgemuth 2009: Ch. 8.), differs from direct and indirect insertion in that loan verbs are accommodated by using complex constructions:

> the borrowed elements remain mostly uninflected and more or less neutral with regard to their part-of-speech membership. The other part of the complex predicate is often a "light verb" which has an auxiliary-like function and bears the inflection or — more generally — all grammatical information of the resulting compound predicate, while the semantic information is by and large associated with the loanword part of the complex verb. (Wohlgemuth 2009: 102)

In short, the light verb strategy combines the deverbal noun of the loan verb with a *light verb*. Light verbs have sparked and continue to spark much interest as a research topic (e.g. Cattell 1984; Algeo 1995; Akimoto & Brinton 1999; Matsumoto 1999; Tanabe 1999; Claridge 2000; Iglesias-Rábade 2001; Anderson 2006; Elenbaas 2013). With a usage frequency of 23.8% in Wohlgemuth's (2009) database, the light verb strategy is cross-linguistically the second most common accommodation strategy,

mainly being used in SOV languages. Similar to indirect insertion, the light verb strategy "involves more integrational effort" (Wohlgemuth 2009: 136) than direct insertion before the loan verb becomes operational. An example of the light verb strategy is shown in (15), where English loan verb *to park* is integrated into Turkish by means of light verb *yapmak*, meaning 'to be'. The *-mak*-marker denotes the infinitive.

(15)    Turkish

*park yap-mak*
park be-INF
'to park'

< English *to park* (Wohlgemuth 2009, example based on Lewis 1985: 155)

Light verbs, albeit cliticised, left-headed or right-headed, most commonly bear the meaning of *do* and *make*, as in (16): Russian *perevesti* is combined in Uzbek with light verb *qilmoq* ('to do').

(16)    Uzbek

*perevesti qilmoq*
translate *do*
'to translate'

< Russian *perevesti* 'to translate' (Wohlgemuth 2009: 107, example based on Schlyter 2003: 162)

In some languages, for example in the Austroasiatic language Gadaba, copulas *be* and *become* can be used as light verbs as well. This is shown in (17), where Gadaba uses *er-* ('to become') to integrate Telugu *pelu* into the language.

(17)    Gadaba

*pel er-*
explode become
'to explode'

< Telugu *pelu* 'to explode' (Wohlgemuth 2009: 110, example based on Bhaskararao 1998: 352–353)

Whereas *do* and *make* are generally used to accommodate transitive loan verbs, *be* and *become* are used to accommodate intransitive loan verbs (Wohlgemuth 2009: 109). The set of light verbs sporadically extends to verbs which are not traditionally seen as semantically 'light' (as defined in Section 7.2.1.). In example (18), for instance, Carib

untraditionally uses *poko*, which means 'busy with', in order to grammatically accommodate *pentiré*, a loan verb from Guianese French Creole.

(18)  Carib

> *pentiré poko man*
> paint busy.with 3SG.COP
> 'he is painting'

> < Guianese French Creole *pentiré* 'to paint' (Wohlgemuth 2009, example based on Renault-Lescure 2004: ex. 19)

The use of light verbs throughout the history of English, and more specifically during the Late Middle English period, will be discussed in great detail in Chapter 7.

### 2.3.1.4. Paradigm insertion

The fourth and last accommodation strategy is *paradigm insertion* (Wohlgemuth 2009: Ch. 9.):

> the loan verb is *not* adapted to the recipient language's morphology at all but is borrowed along with significant parts of the donor language's verbal inflectional morphology. (Wohlgemuth 2009: 118)

Since the loan stem in this strategy is borrowed from its source language, and the inflections are source-language inflections as well, it can prove difficult to distinguish paradigm insertion from single-word code-switching, as defined in Section 5.4.1. (cf. Wohlgemuth 2009: 119–120). As in direct insertion, loan verbs in paradigm insertion are immediately fully operational, meaning that this fourth strategy comes with a low morphosyntactic effort. However, a new inflectional class (i.e. a class containing source-language inflections) needs to be created for the loan, which increases the integrational effort (Wohlgemuth 2009: 135). This may explain why paradigm insertion is extremely rare, with an occurrence rate of only 0.5% and attestations found in only three languages in Wohlgemuth's (2009) database. It is the least frequent accommodation strategy among loan verbs cross-linguistically: paradigm insertion occurs exclusively in intense contact situations characterised by widespread societal bilingualism (Wohlgemuth 2009: 118), such as the French-Middle English contact setting (Section 2.4.3.2.) as described by Ingham (2006, 2009b, 2020) and proponents.

An example of paradigm insertion in the Ajia Varvara variety of Romani is listed in (19). In Ajia Varvara Romani, an Indic language, Turkish loan verbs *okumak* and *yazmak* carry *-sun*, the Turkish 2SG marker, instead of the expected recipient-language *-os*.

(19)    Ajia Varvara Romani

*and o sxoljo ka siklos te okursun ta te jazarsun*
in ART school FUT learn.2 COMP read.2SG and COMP write.2SG
'at school you will learn how to read and write'

< Turkish *okumak* 'to read' and *yazmak* 'to write' (Wohlgemuth 2009: 119, example based on Bakker 2005: 9)

Some cases of paradigm insertion even involve the borrowing of grammatical forms without recipient-language equivalents. An example is provided in (20) for Domari, an Indo-Aryan language. Apart from borrowing Arabic verbs, Domari even borrows their inflections, and for modal verb *xallī* 'to let' even their grammatical gender distinction (*hum* for the masculine third person plural), whereas Domari does not distinguish between grammatical gender of native words.

(20)    Domari

*xallīhum skunnhōšad barariyamma*
let.3PL live.SBJV.3PL outside.LOC
'Let them live outdoors.'

< South Levantine Spoken Arabic *xallī-hum* 3PL.M of *xallī* 'to let' (Wohlgemuth 2009: 120, example based on Matras 2005: 249 ex. 8)

We have encountered instances of paradigm insertion in our dataset for Late Middle English, and will expand on those in Sections 2.4.3.3. and 4.3.2.2.

## 2.3.2. Conventionalisation

Following Backus (2015: 28; also see 2021: 116),

> [b]orrowing is the diachronic process whereby [other-language elements], through their usage, get entrenched in individual speakers, and spread through the speech community as accepted and conventional words in the language.

Indeed, loan words typically enter their recipient language as low-frequency items, and after they have accommodated to their recipient language, they gradually become entrenched or conventionalised (Myers-Scotton 1993a; Treffers-Daller 2005; Matras 2009; Backus 2015), or shared among a group of speakers. This is also the case for the patterns which arise in the process of loan word accommodation, including biases in accommodation as discussed in Chapters 3. and 4.

*Conventionalisation* takes centre stage in Schmid's (2015, 2020) and Schmid & Mantlik's (2015) *Entrenchment-and-Conventionalization Model* (abbreviated EC-Model). It can be defined as "the continuous mutual coordination and matching of communicative knowledge and practices, subject to the exigencies of the

entrenchment processes taking place in individual minds" (Schmid 2015: 10). Words which become conventionalised also gain in frequency. Rohde, Stefanowitsch & Kemmer (2000: 5) consider conventionalisation a critical step for a word to become a loan word, meaning that "it must spread through the speech community in a series of usage events, each of which will affect its integration into the conceptual system of the borrowing language". The term 'conventionalisation' refers to the macro-level, or the level of the speech community and is, therefore, distinct from *entrenchment*, which refers to the micro-level, or the level of individuals (Schmid & Mantlik 2015: 584). However, both phenomena interact with each other (Schmid 2020). Following Backus (2021: 113), better-entrenched forms are easier to activate than less-entrenched forms. Since entrenchment refers to individual cognition, entrenchment levels of (loan or native) words differ from person to person, and can even fluctuate (Backus 2015: 27). The focus of this thesis is on conventionalisation, where we look at the integration of borrowings at the level of the speech community. In Section 5.3.2. we expand on entrenchment, where we look at the level of individual cognition.

Conventionalisation can be interpreted by calculating the frequencies of constructions or aggregate patterns as attested in traditional balanced corpora (Schmid 2016: 14). That is, the higher the frequency of constructions or patterns, the higher the degree of conventionalisation or acceptability (Schmid 2013: 106; Schmid 2016: 122). An example of a corpus purposely built to trace entrenchment (i.e. individual) patterns is the *Early Modern Multiloquent Authors* corpus, abbreviated EMMA (Petré et al. 2019). However, Backus (2015: 27, 34) has argued against the use of corpus frequencies to determine the degree of entrenchment in bilingualism research; instead, he suggests to rely on experimental research, such as acceptability tasks.

## 2.4. Language contact settings

Languages in contact can undergo change, a process which is called *contact-induced language change*, and which will be introduced in Section 2.4.1. The outcome of a given language contact setting or situation — *contact setting* or *situation* in short — is strongly dependent on a number of factors described in the literature (Section 2.4.2.). After discussing those factors, we will present the two contact settings focused on in this thesis. First, we will expand on the French-Middle English contact setting (Section 2.4.3.), which took place in Medieval England between roughly 1100 and 1500. Second, we will present the English-Dutch contact setting (Section 2.4.4.), for which we focus on the current setting in the Low Countries.

## 2.4.1. Contact-induced language change

Research on *contact-induced language* (or linguistic) *change* has been pioneered by Weinreich (1953) and has then been further boosted by Thomason & Kaufman (1991), who have investigated contact from a historical perspective (cf. Matras 2013: 8). Thomason (2001: 62) has defined contact-induced change as "any linguistic change that would have been less likely to occur outside a particular contact situation". It follows that the basic condition for contact-induced change to take place is that two or more languages are in contact (Gardani 2008: 11). Following Matras (2013: 8),

> [m]odels of contact-induced language change [in the past] have taken the position that languages are self-contained systems that influence one another either as a result of the greater social prestige that one language enjoys over another, or else in an attempt by speakers to fill so-called 'gaps' in the lexical and grammatical representation of the recipient language, by extending it to cover functions that are present in the donor [i.e. source] language.

However, Matras (2013: 10) has observed that newer models of contact-induced change (e.g. Matras & Sakel 2007; Matras 2007, 2009) have stopped regarding innovations as 'gaps'. Instead, innovations are considered

> an attempt by speakers to make optimal use of the full range of expressive structures within the linguistic repertoire that is at their disposal and regards individual speakers' creativity in discourse as a major trigger for long-term language change (Matras 2013: 10)

Contact-induced change is nowadays considered a highly common phenomenon across languages (Seifart 2019: 13) — especially when the languages have an asymmetrical relation (Turan et al. 2020) — and it "is manifold in its manifestations" (Seifart 2019: 13). For instance, some authors claim that Middle English has changed drastically due to language contact (e.g. Schendl 2012: 506). However, the literature is not in agreement on what domains of the language are vulnerable to contact-induced change (cf. Turan et al. 2020). Following Romaine (1989: 67), contact-induced change generally does not disrupt genetic links between languages, pidgins and creoles being the exception rather than the rule.

In their paper on constraints on contact-induced change, Heine & Kuteva (2008: 59) have included the following figure to distinguish between the main types of contact-induced linguistic transfer (Figure 1):
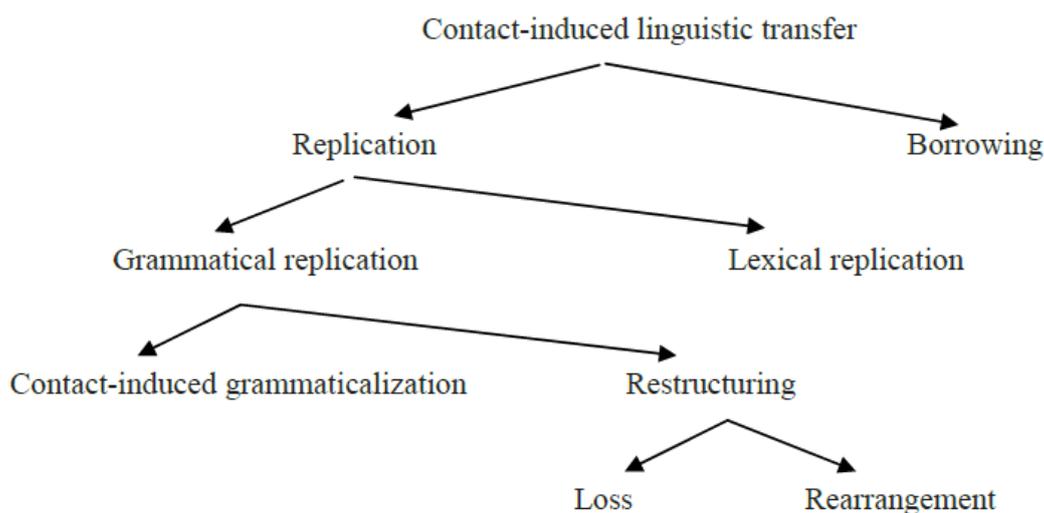
Figure 1: Main types of contact-induced linguistic transfer as illustrated in Heine & Kuteva (2008: 59).

Unless otherwise specified, this thesis will focus on the category of borrowing as a result of contact-induced change, which can be found on the second level in Figure 1.

### 2.4.1.1. Direct versus indirect contact-induced change

The literature has largely focused on cases of contact-induced change caused by *direct transfer*, where "the actual phonemes and/or morphemes (both lexical and grammatical) occur in a language A on the grounds that they have been introduced into A from another language B" (Gardani 2008: 12). However, as we will explain in Section 2.5., this thesis is mainly concerned with cases of contact-induced change caused by *indirect transfer*. To our knowledge, this is a generally neglected aspect where

> one language A rearranges its own inherited phonemic, syntactic and morphological material in such a way that, under the influence of the adjoining foreign model, it moves structurally closer to the neighbouring language and structural convergence results. (Gardani 2008: 12)

Simply put, the recipient language adopts elements of its source language. However, the recipient language can also impact the source language, as will be seen for the case of Anglo French and Middle English (cf. Section 2.4.3.2.). In a similar vein, Thomason (2001: 62; also see Bowern 2013: 422) has proposed that, although "changes are actually motivated by internal pressure within the language", "they would have been less likely to happen if the initial contact-induced change had not happened". In other words, language contact can trigger or accelerate ongoing developments in languages. An example of contact-induced change going beyond direct influence has been found by Trips & Stein (2019) for French loan verbs, since French influx turned out to affect Middle English argument structure.

## 2.4.2. Factors affecting the outcome

Several researchers have stressed the importance of considering the interplay between factors (e.g. Weinreich 1953: 83), since the type of borrowing and the probabilistic outcome of a given language contact setting are dependent on the characteristics of that contact setting. Following Muysken (2013: 710), for instance,

> [i]t is important to realize that there are many different results of language contact. Languages do not interact in a single way, but rather in many different ways, depending on the social setting of the contact.

This can be exemplified by means of the three types of borrowing discussed in Section 2.2.4.: lexical borrowing, syntactic borrowing, and morphological borrowing. Whereas lexical borrowing can occur even in less intense contact settings, syntactic borrowing only occurs when contact is intense and stretches over a longer period (Thomason & Kaufman 1991: 74–75; Bowern 2013: 419). The third subtype, morphological borrowing, is rare, even under intense contact (Thomason 2016: 43). From this example it follows that different circumstances and conditions result in different outcomes, and that instances of language contact should be investigated on a case-by-case basis. Muysken (2013: 710) has referred to this approach as the "scenario approach". Thomason (2001: 77) has claimed that the outcome of contact-induced contact is, in fact, unpredictable; nevertheless, this section will attempt to provide a non-exhaustive list of factors possibly affecting the outcome of contact settings.

The literature has proposed a wide range of notions to characterise contact settings (cf. Muysken 2013: 710). A first set of factors operates language-internally and is concerned with differences between the languages in contact. Such factors have also been called 'intra-linguistic' by Gardani (2008: 45). For example, much is dependent on the (lack of) compatibility (Sankoff 1998) and the *typological distance* (Thomason & Kaufman 1991: 72; Thomason 2001: 76) between source and recipient language. Thomason & Kaufman (1991: 72) have defined compatibility as "a measure of structural similarity that applies to linguistic categories and their combinations, including ordering relations". It is widely accepted that relatedness (or typological proximity) facilitates the transfer of linguistic material (e.g. Meillet 1921; Weinreich 1953: 89; Moravcsik 1975b), even of features characterised as "highly marked" (Thomason 2001: 77). According to Thomason & Kaufman (1991: 72), that is because languages with more complex grammatical systems have more strongly connected categories and will, therefore, have lower correspondence to other languages. Thus, when languages are related, the correspondence between the connected categories in those languages will be higher. However, Bowern (2013: 417) has asserted that "the reason why material and structure are more easily transferred

between closely related languages might just be their typological similarity rather than their relatedness[15]". That is because "the more closely related two languages are when they come into contact, the less time they have had to diverge, develop new patterns and coin new vocabulary" (Bowern 2013: 417). Whereas research on borrowing has shown that relatedness facilitates borrowing, code-switching research has shown that language users who switch between typologically more related languages have to pay a higher switch cost than when they switch between typologically less related languages (Deibel 2020). Another factor which influences the outcome of a given contact setting is the grammatical distance between the languages in the language pair (Deibel 2020): do the grammatical structure and patterns of both languages coincide? If patterns, such as parts of speech, coincide, then the grammatical distance will be smaller, and transfer will be easier (Deibel 2020). Similarly, Deuchar (2005) has found that a greater degree of *congruence* between categories in two languages increases the likelihood that transfer can happen[16]. In her work, congruence is defined as "a notion of equivalence between the grammatical categories or word classes of different languages" (Deuchar 2005: 256). She distinguishes between two types of congruence, namely *paradigmatic* and *syntagmatic* congruence. Whereas paradigmatic congruence is the "similarity or equivalence between the grammatical categories of two languages", syntagmatic congruence relates to the "similarity of word order" (Deuchar 2005: 256). Following Hout & Muysken (1994: 52), the source-language frequency of the element to be borrowed has to be taken into account as well, since frequency is directly proportional to the likelihood that the element will be borrowed.

A second set of factors is of a historical and societal nature, and has also been called 'extra-linguistic' by Gardani (2008: 43). We follow Weinreich (1953: 83), who has claimed that,

> when a group of some size brings two languages into contact, idiosyncrasies in linguistic behavior tend to cancel each other, while socially determined speech habits and processes characteristic of the group as a whole become significant.

This ties in with the more recent idea by Thomason & Kaufman (1991: 35; also see Sankoff 2001; Thomason 2001) that — although linguistic characteristics are important to take into account when investigating the outcome of language contact — "it is the sociolinguistic history of the speakers, and not the structure of their language, that is the primary determinant of the linguistic outcome of language contact". Moreover, Thomason (2001: 11) has claimed that "all aspects of language structure are subject to transfer from one language to another, given the right mix of

---

[15] Additionally, Bowern (2013: 424) has observed that, "because related languages contain material inherited through common descent, the possibility exists that similarities will fail to be identified as due to contact".

[16] Note that this finding in Deuchar's (2005) work originally applies to code-switching, and not to borrowing.

social and linguistic circumstances". According to Matras (2013: 8), by "link[ing] the borrowability of structural categories on a scale with the sociolinguistic and cultural dimension of contact", Thomason & Kaufman (1991) go beyond the traditional hierarchies of borrowability (Section 2.2.2.1.).

The literature proposes various extra-linguistic predictors for the outcome of contact, which relate to the social constellation in which contact takes place: prestige and political dominance of the source language as well as attitude towards that language (Thomason & Kaufman 1991: 72), the kind of interactions taking place (Campbell 1998), and the proportion of source- and recipient-language speakers (Thomason & Kaufman 1991: 72) are only a handful of factors. Other predictors are the mobility of the population, with a mobile population resulting in increased borrowing rates (Bowern et al. 2011), and social and demographic factors, with transfer being facilitated in both directions when the socio-economic status of both groups is similar (Bowern et al. 2011: 2). This thesis specifically focuses on another key factor in contact settings, namely the *intensity of contact* (e.g. Thomason & Kaufman 1991; Campbell 1998; Thomason 2001; Matras 2007; Bowern et al. 2011). The following citation by Thomason & Kaufman (1991: 47–48) explains how this predictor impacts the outcome of a given contact setting:

> Intensity of contact in a borrowing situation crucially involves factors of time and of level of bilingualism. If few speakers of the borrowing language are bilingual in the potential source language, then normally only words will be borrowed […]. However, if there is extensive bilingualism on the part of [recipient-language] speakers, and if this bilingualism persists over a long period of time, then substantial structural borrowing is a probability.

It follows that intensity of contact positively affects borrowing, and that the number (as well as the types) of borrowings increase as intensity increases. That is because — in high-intensity contact settings — contact stretches over a longer period of time, and numerous speakers are bilingual. Factors closely related to intensity of contact are the *duration of contact* (called *time-scale* in Kerswill & Williams 2002) and the *level of bilingualism*, which are directly proportional to intensity. That is because "the longer two languages are in contact, the more time there is for speakers of one or both groups to become bilingual" (Thomason 2001: 66).

## 2.4.3. French-Middle English contact setting

Throughout its history, the English language has been in contact with various languages, for instance with Celtic, but Celtic influence on the English lexicon has traditionally been called negligible (Filppula & Klemola 2010). Scandinavian languages, too, have exerted influence on the English language, and in particular on the northern varieties (e.g. Thomason & Kaufman 1991: 280–282; McWhorter 2002).

However, contact between Old Danish/Old Norse and English (Crespo 2000: 29) is not covered in this thesis, since it took place during the Old English period (roughly 787–1042) and, therefore, predated French influx. Consequently, by the Late Middle English period, most Scandinavian loans had already fully integrated into the English stock. In this thesis, Late Middle English is defined as the period of Middle English ranging from roughly 1350 to 1500.

### 2.4.3.1. Historical context

Following Timofeeva & Ingham (2018: 197) in their review article, "there is a clear need in […] historical research for scholars to reinvestigate earlier stages of English as a contact language". Additionally, Rothwell (1980: 143) has stressed that "the whole question of lexical borrowing in the Middle Ages must be tackled in a multilingual context". This section will, therefore, describe the French-Middle English contact setting by focusing on the mutual relationship between Middle English and French. However, we must not forget that Latin was another important contact language in Medieval England, and that the linguistic system has even been called "trilingual" (Crespo 2000: 23).

The French-Middle English contact setting is well-known and particularly widely-documented, and sparks the interest of many researchers until today (see e.g. Jespersen 1905; Rothwell 1976; Dekeyser 1986; Thomason 2001; Culpeper 2005; Mugglestone 2006; Brinton & Arnovick 2011; Ingham 2012; Baugh & Cable 2013; Durkin 2015). Its appeal does not come as a surprise, since the 1066 Battle of Hastings, in which William of Normandy (also called William the Conqueror) won over Medieval England, is considered "perhaps the single most important event affecting the linguistic development of English" (Brinton & Arnovick 2011: 241). It is, therefore, not without reason that the Norman Conquest marks the end of the 'Old English' era and the beginning of the 'Middle English' era (Emonds & Faarlund 2014), a period generally characterised by great linguistic change (Denison 2003: 68; as proven quantitatively by Nevalainen et al. 2020: 26). This linguistic change has been characterised as far greater than between the Old English period and the Late Modern English period (Denison 2003: 68). According to Nevalainen et al. (2020), this ties in with Dixon's (1992: 67) punctuated equilibrium model, according to which long periods of linguistic stability (hence 'equilibrium') alternate with periods of major linguistic change.

### 2.4.3.1.1. Two stages

Contact between French and Middle English, a Romance and Germanic language respectively, can be localised in Medieval England, where Norman French entered the isles after having been conquered by the Normans in 1066, and contact between the two languages proceeded until roughly 1500. It has to be noted, however, that

English had already adopted some loans from French before the Norman Conquest, due to its military relations with France and France's cultural prestige at the time (Strang 1970). The contact setting under investigation can best be described in two stages, during which the intensity of contact and the number of speakers of both English and French changed.

The first stage of contact was kickstarted after a couple of decades without mutual intelligibility between speakers of English and French. According to Bartlett (2000: 1), "[a] small armed group speaking a language incomprehensible to the majority of the population controlled virtually all landed wealth". Whereas the Norman aristocracy spoke French, the English peasantry spoke English (e.g. Kibbee 1991: 10), and the Norman elite ruled from afar (McWhorter 2002: 253). Therefore, "the impact of the French language was limited to a small (albeit highly influential) portion of the population" (Kibbee 1991: 11). This changed when the Norman elite, which had settled in Medieval England, started learning English as a second language (Mugglestone 2006). Also, the Norman French language spoken by the Norman elite became influenced by English and gradually developed into *Anglo French*. This superstrate local contact variety of French started being used as a high-prestige written governmental and administrative language alongside Latin (Coetsem 1988; Dalton-Puffer 1996; Culpeper 2005; Short 2007; Matras 2009; Fischer, De Smet & Wurff 2017). Until 1250, borrowing rates from French remained generally low and the borrowings at hand were mainly limited to cultural borrowings, relating to domains such as religion and nobility (Baugh & Cable 2013). Those earlier borrowings were mostly based on Norman French, the mother tongue of the Norman aristocracy in Medieval England, and they typically followed the English stress pattern, where the stress is on the first syllable, as in *móral* (Brinton & Arnovick 2011: 249). Other examples of Norman French borrowings are *champion*, pronounced with a /tʃ/ sound, and *gentle*, pronounced with a /dʒ/ sound (Brinton & Arnovick 2011: 249).

In the second stage of contact, which started roughly from the 13th century onwards, the number of native Anglo-French speakers gradually declined, since the Norman elite stopped travelling between Normandy and England (Ingham 2012: 160–161). Also, English started taking over again from French as a written administrative and governmental language. In fact, "while *some* English speakers learned French […], virtually *all* Normans and Angevins[17] became bilingual in English within no more than 250 years of the Conquest" (Thomason & Kaufman 1991: 308). Despite the decline of the use of Anglo French, in the 13th and 14th centuries it surged as a language of record-keeping (Ingham 2012: 27) during the 13th-century "information explosion"

---

[17] The Angevins were a "royal house of England, which reigned from 1154 to 1485" (Encyclopædia Britannica n.d). The Angevins were of French descent.

(Hunt 2008: 151), and it started being taught to speakers of English (Kibbee 1991: 1). This may explain why the borrowing rates had never been higher than during this period, when they surged to roughly 30% (Dalton-Puffer 1996: 12). This was true in particular for the second half of the 14th century (Jespersen 1905; Dekeyser 1986; Mugglestone 2006; Baugh & Cable 2013), the so-called *borrowing peak* (Finkenstaedt & Wolff 1973; Dalton-Puffer 1996). Following Bowern et al. (2011), a borrowing rate of 30% is considered high, as it is three times higher than the cut-off point of 10%. Additionally, older Germanic words were increasingly replaced by French loans, a process which Dalton-Puffer (1996) has referred to as the *Wortschwund* (lit. 'word loss'). Whereas borrowings before 1250 had largely been limited to the cultural domain, borrowings after 1250 belonged to a much wider variety of domains (Baugh & Cable 2013). Those later borrowings mainly originated from Central French[18], the high-prestige but foreign variety of the Ile-de-France. Examples of such borrowings are *chandelier* and *genre*, pronounced with /ʃ/ and /ʒ/ sounds respectively (Brinton & Arnovick 2011: 251). Central French borrowings typically followed the French stress pattern, in which the stress is on the last syllable, as in *morále* (Brinton & Arnovick 2011: 252).

The decline of French as a contact language in Medieval England started in the late 14th century, when French moved from being an actively spoken language to a language taught in school settings (Ingham 2012: 35). This demise has repeatedly been linked to the Statute of 1362, which stated that "French no longer be the language of government and of the legal system, because that language was too poorly understood" (Kibbee 1991: 58). However, Kibbee (1991: 58) has also added that — ironically speaking — the statute was written in French. Other researchers have deemed it unlikely that the Statute of 1362 has marked the end of massive influx from French, since French continued to be used as before (e.g. Rothwell 2001: 539). Other explanations offered for the demise of French have been the 1337–1453 Hundred Years' War (e.g. Schendl 2012: 507–508) and the 1349–1362 Black Death pandemic. Whereas the Hundred Years' War has frequently been rejected as the cause of the decline, the Black Death sounds more plausible to most researchers (cf. Ingham 2012: 32, 35). That is, the French-speaking clergy took care of the patients during the outbreak and were, therefore, at higher risk of death. Since the clergy also provided French-speaking education in schools, the transmission of French was disrupted due to those increased death rates (Ingham 2012: 35). As of the 15th and 16th centuries onwards, Medieval England became largely monolingual (Crespo 2000: 25).

---

[18] However, there is some disagreement as to whether or when Anglo French was superseded as a contact language by Central French (see e.g. Rothwell 1998; Schendl & Wright 2011; Miller 2012).

More remarkable is that many of the borrowings from French have stood the test of time and have found their way into Present-day English, 28% of the current vocabulary being of French descent (Finkenstaedt & Wolff 1973).

### 2.4.3.1.2. *Anglo French*

The source language in the contact setting in Medieval England was Anglo French. However, in this thesis Anglo French, as well as the other historical varieties of French observed at the time, will be referred to as 'French'. More concretely, 'French' will be used as an umbrella term for the following historical varieties: Central French (also called Continental French), Norman French, and Anglo French (also called Anglo Norman and Anglo-Norman French). Researchers generally agree that Central French was the variety of French spoken in Mainland France (Paris), and Norman French the variety of French spoken in Normandy. However, the definition of Anglo French is a topic of discussion. For example, whereas 'Anglo French' and 'Anglo Norman' are often used interchangeably, they are not used interchangeably by all researchers. Wogan-Browne (2009: 1), for example, has suggested that the term Anglo Norman "generally denotes French texts composed in the British Isles from the Conquest to the early fourteenth century", whereas Anglo French "refers to textual imports from the continent into England and to contacts between England and the continent in the later fourteenth and early fifteenth centuries". According to those definitions, both terms are used to denote different stages of contact (Wogan-Browne 2009: 1). In this thesis Anglo French and Anglo Norman will be considered synonyms for the insular variety of Norman French in Medieval England during the period of contact.

Much has been said about the relationship between Central French and Anglo French[19]. Wherever possible, dictionaries such as the OED (Oxford University Press 2009) and MED (Lewis 1952–2001) distinguish between forms of both varieties, yet there exists no common consensus as to how Anglo French relates to Central French. For instance, researchers have investigated whether Anglo French is an independent language, a variety, or a dialect of Central French (cf. Trotter 2003: 43). Trotter (2003: 43) has concluded that Anglo French is a variety, and although it is certainly insular, this does not imply that it is isolated. Traditionally, many researchers have claimed that Anglo French and Central French differ fundamentally, since Anglo French was not a language spoken by native speakers in the decades after the Norman Conquest, and it was isolated from Central French (e.g. Kibbee 1991). However, according to Ingham (2009a: 44), this does not mean that Anglo French

---

[19] However, Schauwecker (2019: §44) has found that the use of Anglo French differs widely among individual authors.

"can be treated as a foreign language barely understood by its users [i.e. inhabitants of Medieval England]".

Ingham (2009a: 44) has acknowledged some differences between Anglo French and Central French: for example, Anglo French uses *à les* for the combination of preposition *à* ('to') and plural definite article *les*, whereas Central French uses contraction *aux*. According to Schauwecker (2018), such factors which distinguish Anglo French from Central French showed close parallels to the English language as used at the time (cf. examples (23) and (24) in Section 2.4.3.2.). There also existed phonological differences between Anglo French and Central French (e.g. Ingham 2012: 68). Then again, Anglo French initially strongly resembled Central French since it underwent similar developments as Central French (Ingham 2006, 2009a, 2011b). For instance, Anglo French and Central French simultaneously replaced *nul* in negative clauses by indefinite *aucun* (Ingham 2011b). The similarities between the varieties started being interrupted in the second part of the 14th century, when the inhabitants of Medieval England started to move away from using French (Ingham 2006: 25).

### 2.4.3.1.3. *Historical code-switching*

The Medieval English linguistic climate was particularly tolerant to other languages, as can be seen from the frequent *code-switching* between Middle English and Anglo French, but also Medieval Latin[20]. An instance of a historical code-switch between French and Middle English is provided in (21), where all French-origin material is printed in bold.

(21)  **Il jecte un graund** *brickbat* **que** *narrowly mist*

'He threw a large piece of brick that narrowly missed.' (Chambers 1932: lxxxii in Schendl & Wright 2011: 16)

Code-switching in Medieval England has often been referred to under the term 'macaronic writing' (e.g. Wright 1992), or a "comic Latin verse form characterized by the introduction of vernacular words with appropriate but absurd Latin endings" (Encyclopædia Britannica n.d.). Due to its comic and pejorative connotation, this term will be avoided throughout this thesis.

Switching in Medieval England was "extraordinarily common"[21] (Wright 2011: 101; also see 1992, 1995, 2010; Schendl 2000, 2004; Ingham 2011a; Schendl & Wright 2011), and it occurred in specialised text genres, such as business writing and official letters. The different ways in which code-switching manifested itself

---

[20] Since code-switching happened over a large period, Wright (2011) has distinguished three broader stages; however, we will not expand on those.

[21] However, this was the case in the higher ranks. The lower classes were largely monolingual (Schendl 2000).

(e.g. English words unmodified in Latin texts, English words dressed up as French in French texts, French words dressed up as Latin in Latin texts) have been discussed in detail by Rothwell (2000: 216–231).

A notable example of historical code-switching is the 1403 letter from Dean Richard Kingston to King Henry IV (Schendl 2000: 81), depicted in (22). Richard Kingston constantly switches between English and French, and includes both switches within and across sentences (cf. intra- and intersentential switches in Section 5.4.1.). The switches persist over longer stretches of text, and all French-origin material in example (22) appears in bold.

(22)   *Please* **a vostre tresgraciouse Seignourie entendre que a-jourduy apres** *noone* … **qu'ils furent venuz deinz nostre** *countie* **pluis de .cccc. des les rebelz de Owyne, Glyn, Talgard, et pluseours autres rebelz des voz marches de Galys** … *Warfore, for goddesake, thinketh on your beste frende, god, and thanke hym as he hath deserued to yowe! And leueth nought that ye ne come for no man that may counsaille yowe the contrarie* … **Tresexcellent, trespuissant, et tresredouté Seignour, autrement say a present nieez. Jeo prie a la benoit trinité que vous ottroie bone vie ove tresentier sauntee a treslonge durré**, *and sende yowe sone to ows in help and prosperitee; for in god fey, I hope to almighty god that, yef ye come your owne persone, ye shulle haue the victorie of alle youre enemyes* … **Escript a Hereford, en tresgraunte haste, a trois de la** *clocke* **apres** *noone,* **le tierce jour de Septembre.**

'May your most gracious Lordship be pleased to hear that today, in the afternoon … more than 400 of Owen, Glyn and Talyard's rebels, and several other rebels from your Welsh borders, entered our country. Wherefore, for God's sake, set your mind on God as your best friend, and thank him for the favours he has bestowed upon you. And do not for any reason fail to come, whatever advice to the country you may receive from anyone … Most excellent, most powerful and most redoubtable Lord, let me be denied / refused in some other way! I pray the blessed Trinity that you be granted good life with perfect health for a long time to come, and may [the Trinity] send you to us soon in help and prosperity; for I faithfully pray to almighty God that, if you yourself come in person, you will be victorious over all your enemies … Written at Hereford in the utmost haste at three o'clock in the afternoon on the third day of September.' (translation reproduced from Gardner-Chloros 2009: 40)

This letter is proof of the degree of "basic social acceptability" of code-switching (Schendl 2000: 81), as it is even used in a letter addressed to the king.

Despite instances of frequent code-switching, opinions are generally mixed as to how strong the influence of French on Middle English actually was.

### 2.4.3.2. Intensity of contact

According to Kibbee (1991: 1),

> the linguistic relationships in England in the Middle Ages have already been treated in a number of works, but never in a comprehensive manner and all too often by authors who seem to have a professional or emotional stake in under- or over-estimating the domination of French.

In short, whereas contact between the French and Middle English languages is generally uncontroversial, the intensity of contact — as well as the demise of French (Section 2.4.3.1.) — is a hotly debated topic among contact researchers. According to Kemenade (1987: 2–3), for instance, "it is difficult to assess the extent of foreign influence on English by the […] French invasions in the [Old English] and early Middle English […] period". Also, French and English were used for different purposes: whereas French — like Latin — was more common in the formal and official register as well as in written language, English was more common in informal and colloquial settings as well as in spoken language (e.g. Crespo 2000: 24).

Although contact between French and Middle English spanned more than four centuries, the literature traditionally assumes that the overall intensity of contact stayed relatively weak. Important proponents of this view are Thomason & Kaufman (1991: 308). Their word-for-word argumentation in favour of the low intensity of contact is cited below (Thomason & Kaufman 1991: 308):

"1. There were never many speakers of French in England.

2. They began giving up French by 1235 at the latest.

3. There is no reason to suppose that any large proportion of native English learned French between 1066 and 1250; after that point they had no reason to do so.

4. Dialects of English most in contact with French underwent no simplification that can be traced to French; they are among the most conservative Middle English dialects, and no doubt would have been so with or without contact with French.

5. Simplifying traits in Standard English are imported from the East Midlands; sometimes these traits originated in the North.

6. The massive French influence on English vocabulary, *followed* by the mild influence on English morphology and syntax, and the practically trivial influences on English phonology, took place at a time when there were practically no competent French speakers around for an Englishman to talk to."

Similar to Thomason & Kaufman (1991: 308), proponents of this view have proposed that French has widely exerted lexical influence on Middle English, *chandelier* and *confection* only being a few out of many examples. A classic example is the use of French loan words for prepared meat (e.g. *mutton*, *beef*, *pork*), whereas the animal names were referred to by means of native English words (e.g. *sheep*, *cow*, *swine*) (Brinton & Arnovick 2011: 249). Although French influence may have extended to orthography and style (such as metre and rhyme in poetry), most researchers have suggested that contact between French and Middle English was not particularly intense. Therefore, they have been sceptical about influence having extended to syntax as well, or if it has, they have argued that the influence has remained modest. For example, in her seminal work on the history of English, Mugglestone (2006) has not mentioned the syntactic influence of French on English at all. Furthermore, Fischer (2013: 40; also see Fischer, De Smet & Wurff 2017: Ch. 4.) has asserted that "whenever a voice supporting French influence on syntax is raised, the arguments are usually not fully persuasive, however much effort may have been put into proving the case". Other researchers claiming that syntactic influence of French on the Middle English language was unlikely or highly limited are, among others, Lass (1987: 54–61), Rot (1991), Burnley (1992), McWhorter (2002: 251), Townend (2006), Miller (2012), and Baugh & Cable (2013: 162).

Bearing in mind that borrowing is by no means restricted to lexical items (cf. Campbell 1998), a more recent school of thought has re-evaluated the traditional claim. The current suggestion is that syntactic influence in Medieval England has traditionally been misjudged, and that the presence of French has had more impact than traditionally acknowledged. An important proponent of this view is Rothwell (1968; also see 1976, 1998, 2000, 2001), according to whom French was taught in Medieval England before it was even taught in France, which stresses how influential the French language has been on the isles. Another influential proponent is Ingham (2020: 452; also see 2006, 2009b), who has stated that "the initially monolingual French-speaking and English-speaking communities had largely given way to a bilingual [or diglossic] speech community, at least among higher-status and/or educated individuals". For instance, following Short (1980), Norman and English people frequently married each other and, therefore, learnt each other's languages. Authors such as Rothwell, Ingham, and Haeberli (2010) paint a highly different picture of the contact setting than Thomason & Kaufman (1991: 307–310) and proponents. The argument that contact was strong is logically externalised in a wide variety of examples of syntactic influence. Those range from calqued phrasal constructions (e.g. *s'arrêter court* > *to stop short*) (Prins 1948: 29) to the rise of the *be going to* construction in English (Stein & Trips 2012), and the reflexive use of Middle English verbs (Trips 2020a). The syntactic change "between English and French

which arose after the Norman Conquest" has also been extensively investigated in the *Borrowing of Argument Structure in Contact Situations (BASICS)* project (Stein & Trips 2016). The members of this project have found that contact between Anglo French and Middle English has affected argument structure in English (e.g. Trips & Stein 2019; Elter 2020; Trips 2020b). Related to argument structure, Ingham (2020: 447) has found effects for patient-liability, where "the Patient argument can appear either as subject or direct object without alteration of the verb form", and he gives the English examples of *it* **broke** and *you* **broke** *it* (Ingham 2020: 447). He has concluded that patient-liability in English was driven by contact with Old French, where argument structure liability was common. Other authors as well have made various attempts to explain changes in Middle English syntax by invoking contact with French. Potential developments include the rise of the clausal (or verbal) gerund (Jack 1988), the loss of verb-second (Haeberli 2010), and the development of *wh*-relatives (Mustanoja 1960: 192).

Some researchers, such as Schauwecker (2018) and Schauwecker & Trips (2018), have even suggested that contact between Anglo French and Middle English was reciprocal, meaning that Middle English has influenced Anglo French as well. For instance, in both Middle English and Anglo French, the goal of motion in the directed motion construction "is usually expressed outside the verb by adverbials and particles" (Schauwecker & Trips 2018: 2), meaning that constructions are 'satellite-framed'. This distinguishes Anglo French from other varieties of French, where the goal of motion "is expressed by other syntactic means, i.e. adverbial subclauses (e.g. gerundive constructions)" (Schauwecker & Trips 2018: 2), and where directed motion constructions are 'verb-framed'. Example (23) demonstrates the use of a satellite-framed goal of motion construction in Anglo French, whereas example (24) demonstrates the use of a verb-framed construction typical of the other varieties of French.

(23)    *Le chevalier* **chevaucha** *à Paris.*

        'The knight rode to Paris.' (Schauwecker & Trips 2018: 2, emphasis added)

(24)    *Le chevalier allait à Paris* **en chevauchant**.

        'The knight came riding to Paris. (lit. The knight went to Paris (by) riding.)' (Schauwecker & Trips 2018: 2, emphasis added)

This divergence between Anglo French and Central French may be a result of Middle English exerting influence on Anglo French.

Bailey & Maroldt (1977; also see Domingue 1977) have gone further in saying that Middle English was in fact a French-based creole. However, most authors

(e.g. Thomason & Kaufman 1991: 306–315) have rejected this scenario altogether, since inflectional loss had already started before French started influencing English. One should note here that, according to Stein & Trips (2012: 236), the definition of 'creolisation' provided by Bailey & Maroldt (1977) may differ from the standard definition in that, according to Bailey & Maroldt (1977), "ME [Middle English] is the result of a massive importation of English lexical items into the OF [Old French] spoken at that time on the isle by the upper classes of England".

Regardless of the debate on the extent of influence of French, linguists such as Rothwell (1983: 259–260) have observed that French has mainly influenced the English language in the southwestern varieties of Medieval England, and less so in the northeast.

### 2.4.3.3. Loan verb accommodation strategies applied

French loans in Middle English had to be accommodated to the inflectional patterns and the syntactic paradigms of their recipient language. To this end, the most commonly used accommodation strategy by far was direct insertion (cf. Wohlgemuth 2009). In this thesis the term 'direct insertion' is not only used to refer to verbs, but also more generally to other parts of speech, such as adjectives. The mechanism behind the strategy stays the same for the other parts of speech, namely recipient-language inflections are added directly onto the loan stem (cf. Section 4.1.). Examples of French loan verbs and adjectives integrated in Middle English by means of direct insertion are presented below in (25)–(30).

(25)   *þe ryche me **raymeþ** wiþ-outen eny ryht*

   'The rich robbed me without having the right.' (*Song of the Husbandman*, 1300, HC)

(26)   *And God yeveth and departeth to other folk prosperites and adversites **imedled** to-hepe aftir the qualite of hir corages*

   'And God gives and leaves other people behind with wealth and misfortunes mixed together based on the quality of their courage.' (*Boethius*, 1380, HC)

(27)   *Nichol Brembre wyth his vpherers **purposed** hym the yere next after Iohn Northampton mair of the same Citee*

   'Nicholas Brembre with his defenders intended to become mayor of the same city the year after John Northampton.' (*Appeal(s)*, 1384–1425, HC)

(28) *than is every smal divisioun in a signe departed by two degrees and two, I mene degrees* **conten<u>yng</u>** *60 mynutes.*

'Then is every small division in a sign divided by two degrees and two [minutes], I mean degrees containing 60 minutes.' (*A Treatise on the Astrolabe*, 1391, HC)

(29) *þere ben manye* **periolous<u>e</u>** *passages with outen fayle*

'There are many perilous pathways without doubt.' (*Mandeville's Travels*, c1400, PPCME2)

(30) *And þere nygh is the Foss of Mennon þat is all* **round**

'And there nearby is the Foss of Memnon, which is all round.' (*Mandeville's Travels*, c1400, PPCME2)

In example (25), French loan verb *reimen* ('to plunder, rob') in the present third person singular is inflected by means of an English -(*è*)þ-inflection: *raymeþ*. In example (26), French-origin verb *medlen* ('to mix') is used as a past participle, characterised by the Middle English circumfix *i*⟩…⟨(*è*)*d*. Example (27), then, demonstrates how the verb *purposen* ('to intend') is used in Middle English in the past third person singular as *purposed*. In example (28), French-origin verb *contenen* ('to contain') is used as a present participle, typically ending in *-yng* (including spelling variants). Example (29), next, shows how French-origin adjective *perilous* is used in attributive syntactic position in Middle English, and how it receives -*e*-inflection. In example (30), last, French loan adjective *round* is used in predicative syntactic position, and attributes characteristics to its subject, *Foss of Mennon*.

Although direct insertion is the most commonly used accommodation strategy for French loans in Middle English, the dataset used for Chapter 4. also contains a few instances of paradigm insertion (cf. Wohlgemuth 2009), as in (31).

(31) *in þat chapell syngen prestes* **yndyene<u>s</u>**

'In that chapel sing priests of India (lit. Indian priests).' (*Mandeville's Travels*, c1400, PPCME2)

*Yndyenes* ('Indian') — in this case an adjective — retains its French -*s*-inflection even in Middle English. Such cases of paradigm insertion will be further discussed in Section 4.3.2.2.

## 2.4.4. English-Dutch contact setting

### *2.4.4.1. Historical context*

The presence of English as a language in the world is undeniable. According to the *Eurobarometer* (European Union 2006), 68% of Europeans find English the most useful language to learn apart from their mother tongue. Not surprisingly, in Swaan's (2001, 2002) World Language System, English is currently considered the one and only "hypercentral" language, or a language which holds the language constellation together and which is "the pivot of the world language system" (Swaan 2001: 6). However, note that English has only been holding this position since the 1950s (Swaan 2001: 6). Spanish, Hindi, and Arab are categorised at a level below, namely as "supercentral" languages, which means that they have a large number of native speakers and learners, but not as many as English. Even the widespread diffusion of Mandarin Chinese and — earlier in history — of Latin cannot rival the position occupied by English today (Zenner 2017: 231). Being a *lingua franca* in different registers of written and spoken language, and in various domains such as pop culture, trade and science, English "is a potentially relevant contact language for all the world's languages" (Mair 2019: 15). English elements in other languages (Onysko 2007: 10) — regardless of the recipient language — have been referred to in the literature as *Anglicisms*, although definitions vary across studies (cf. Onysko 2007: 10). MacKenzie (2012: 33–35) has even discussed the presence of *false Anglicisms* (also called pseudo-Anglicisms), or

> coinages that resemble words from the 'prestige' language, English, but which would not be recognized or understood by monolingual English native speakers, and which, if translated from a source text into English by a native speaker, would be substituted by a genuine English word.

Examples from French and Italian, but which are also attested in other languages, are *basket* for *basketball*, *flipper* for *pinball machine*, and *happy end* for *happy ending* (MacKenzie 2012: 33).

Like in many other countries around the world, English is omnipresent in the Netherlands and Flanders, the Dutch-speaking part of Belgium. Both regions, also called 'the Low Countries', have Dutch as their official language. Dutch "is neither a big nor a small language" worldwide (Swaan 2016: 13), and it is considered a mid-size language, since it is one of the forty most spoken languages in the world (De Sutter 2017: 22–23). However, according to Swaan (2016: 12) native speakers of Dutch largely underestimate the importance of their language in a global context, and the Q-value of Dutch — a measure of use value — is higher than its speakers tend to believe.

The influence of English on the Dutch language in the Low Countries is relatively recent, since it has started around the second half of the 20th century and it is ongoing until today[22]. Until the 19th century Dutch mainly borrowed from French[23], which was the prestige language (Zenner 2017: 232), but as of approximately 1950 English started entering the Dutch language. Contact between English and Dutch was sparked by the aftermath of World War II, when the USA — as Europe's 'liberator' — started being idolised, and the English language gained more and more appeal (Sijs 1996; Smans 2011; Zenner 2017). Other factors, such as the Anglo-Saxon orientation of the IT revolution as well as the Americanisation of pop culture and general globalisation, have only accelerated English influence in the Low Countries (Zenner & Geeraerts 2015). Contact between English and Dutch in the Low Countries is similar to that in other (West) European countries, such as Germany (e.g. Stefanowitsch 2002; Onysko 2007; Androutsopoulos 2013; Edwards & Fuchs 2018), Denmark (e.g. Hansen & Lund 1994; Gottlieb 2012), and Norway (e.g. Stene 1945; Graedler 2012, 2014: 20; Andersen 2017; Andersen & Graedler 2020).

## 2.4.4.2. Intensity of contact

In 2012, Sijs has estimated 4% of the Dutch lexicon to be of English origin[24]. Following Bowern et al. (2011), this borrowing rate is considered low, since it is lower than 10%. However, it must be added that this estimation dates back from ten years ago, and that the rate — due to the ongoing and increasing influence of English in the Low Countries (Zenner 2013: 75) — has undoubtedly augmented since then. According to the *Eurobarometer* (European Union 2006), 94% of Dutch people and 91% of Belgian people think knowing other languages than their mother tongue is useful. Apart from Dutch as a mother tongue, the most widely-known language in both the Netherlands and Flanders is English. In the Netherlands, the rate of speakers of English is 87%, and this is considerably higher than in Belgium (59%). This difference between the Netherlands and Belgium is presumably due to the fact that Belgium has several national languages (i.e. Dutch, French, and German), and English in Belgium has competition from those languages.

In any case, contrary to popular perceptions, the intensity of contact between English and Dutch in the Low Countries has always remained weak for three main reasons. First, although most speakers in the Low Countries master English (Stern 1977; European Union 2006), and English has positive connotations and social prestige, it

---

[22] Due to seafaring and the Industrial Revolution it is plausible that English reached the Low Countries before 1950. However, noticeable influence is rather recent.

[23] French influence in Dutch will not be expanded on in this thesis.

[24] According to Janssens & Marynissen (2008: 252), this was 16% in 2000. However, they have not given any references, nor have they described what data this percentage is based on.

is by no means an official language used between speakers of Dutch on a daily basis (Swaan 2016; Zenner & Van De Mieroop 2017: 77). Rather, contact with English is remote and indirect, taking place through the influx of the Anglo-Saxon mass media, including radio, television, the Internet, and advertising, which find their way into the Low Countries in their original language (Booij 2001; Onysko 2007: 44; Androutsopoulos 2013 for English influence on German). This is similar to many other contact settings involving English (Thomason 2001). Second, contact is said to be weak because of the asymmetrical relation between Dutch and English, as it is highly unlikely that the contact is reciprocal (Doğruöz & Zenner 2013; Zenner & Van De Mieroop 2017). Third, according to the *Eurobarometer* (European Union 2006), the community in the Low Countries remains largely monolingual to weakly bilingual; therefore, the contact effect is mainly limited to the lexicon (Zenner, Speelman & Geeraerts 2015). Additionally, many speakers of Dutch "pick up repeated English expressions through the internet and television, but do not often create completely new English sequences on the spot" (Zenner & Geeraerts 2015: 268; also see Stefanowitsch 2002; Onysko 2007: 80). This phenomenon is illustrated in example (32) below, where *when u ask* and *they wanna see* are most likely expressions acquired through memes on social media, whereas *replyen* and *excited* are most likely single-word lexical loans. All English-origin material in the example appears in bold.

(32)  **when u ask** *[NAMES]* **if they wanna see ur** *kersttrui en ze* **replyen** *kapot* **excited**

'When you ask [NAMES] if they want to see your Christmas jumper and they reply very excitedly.' (Twitter, 6 December 2020)

Following Zenner, Speelman & Geeraerts (2012), loan words in weak contact settings as the one in the Low Countries are typically much debated. Organisations which aim to combat against English influence and loan words in the Low Countries are Stichting Nederlands ('Foundation Dutch'), Het Ampzing Genootschap ('The Ampzing Association'), and Bond tegen Leenwoorden ('Federation against Loan words') (Smans 2011: 14–15). That said, a cluster analysis of attitudes towards English in the Netherlands has revealed that the majority of the population tends to be positive towards the English language and does not consider its presence a threat to Dutch (Edwards & Fuchs 2018). Additionally, respondents with more negative attitudes towards English turned out to be generally older, less urban, less educated, and less proficient in English than respondents with more positive attitudes. In sum, although English in the Low Countries is sometimes considered a threat to the Dutch language, Janssens & Marynissen (2008: 252) have not shared this concern, and have suggested that English will not lead to the demise of Dutch. In line with this, most inhabitants of the Low Countries hold rather positive attitudes towards English.

### 2.4.4.3. Loan verb accommodation strategies applied

English loan words in Dutch can typically be accommodated by means of direct insertion (cf. Wohlgemuth 2009), instances of which are given in (33)–(38). This will also become clear from Chapters 3. and 4.

(33)   *heb jij die gedichtjes al in mijn bus* **gedrop***t*?

‘Have you dropped those-those poems in my mailbox yet?’ (CGN)

(34)   *Nederland* **promoot** *het gebruik van de fiets in Zuid-Afrika.*

‘The Netherlands is promoting the use of bikes in South Africa.’ (CGN)

(35)   *en dan kun je je apparatuur* **test***en* *uh of die uh of die nog juist is.*

‘And then you can test your equipment [to check], um, whether it, um, whether it is still all right.’ (CGN)

(36)   *dat* **pop***te* *nu binnen.*

‘It has popped in just now.’ (CGN)

(37)   *tot voor enkele maanden vormde het huis met de één hectare grote wildernis eromheen een oase tussen* **clean***e* *miniparkjes*

‘Until a few months ago the house with the one hectare of wilderness had been an oasis in between clean mini parks.’ (CGN)

(38)   *ik ben nogal* **close** *met de mensen van de gemeente* .

‘I am quite close to the people in town.’ (CGN)

English-origin verb *to drop* is borrowed in Dutch as *droppen*, which is used in example (33) as a past participle. The past participle in Dutch consists of circumfix *ge*⟩...⟨*t*, hence *gedropt*. In example (34), the verb *to promote* is used as *promoot* in the present third person singular. Example (35) presents loan verb *to test* which is used in its infinitival form in Dutch, characterised by the *-en*-marker. Example (36), next, illustrates the use of *to pop* in the past third person singular, marked by *-te*, hence *pop-te*. That *poppen* is used with Dutch-origin adverb *binnen* (‘in’), hence *binnenpoppen* (‘to pop in’), indicates that the loan verb is used as a separable verb, and that it must be well-integrated. In example (37), English-origin adjective *clean* is used in attributive syntactic position, and with *-e*-inflection, inflecting *miniparkjes* (‘mini parks’). Example (38), last, shows how English loan adjective *close* is used in predicative syntactic position, referring back to subject *ik* (‘I’).

## 2.4.5. Comparing the two contact settings

Whereas the two contact settings (and corresponding language pairs) under investigation are fairly similar in some aspects, they differ in others. A non-exhaustive list of the most critical similarities and differences is discussed below.

As for their similarities, not only do both contact settings involve structurally similar recipient languages, namely Dutch and (an older variant of) English, they also involve source languages which are genetically related to their recipient language. In the case of the English-Dutch contact setting, English and Dutch are two closely related West-Germanic languages; in the case of the French-Middle English setting, French is a Romance language and English a West-Germanic language, and both belong to the European branch of the Indo-European languages. This relatedness causes a decrease of the typological (Thomason & Kaufman 1991) and grammatical distance (Deibel 2020) between the source and recipient languages, which is said to facilitate borrowing (e.g. Meillet 1921; Weinreich 1953; Moravcsik 1975b). Also, the languages in both language pairs are relatively congruent, meaning that there are similarities in terms of grammatical categories and word order (cf. Deuchar 2005). However, since English and Dutch are more tightly related than French and English, borrowing between English and Dutch should be a bit easier than between French and Middle English. Interestingly, the borrowing rates for both contact settings (as discussed in the next paragraph) reveal the opposite effect, namely that borrowing rates are higher in Middle English than in Dutch. This may relate to the historical and societal parameters involved (cf. Thomason & Kaufman 1991). A last similarity between the contact settings is that they are both characterised by superstrate influence as opposed to substrate influence.

As for their differences, a first important one is the duration of contact: whereas French has influenced Middle English for no less than four centuries, contact between English and Dutch has only started roughly seventy years ago and is ongoing. Based on the difference in duration of contact, borrowing is predicted to be more likely in the French-Middle English contact setting than in the English-Dutch setting. A second main difference linked to duration is the intensity of contact (e.g. Thomason & Kaufman 1991). Contact between English and Dutch is mostly mediated by popular media and the use of English as a lingua franca in international settings. Following Bowern et al. (2011), borrowing rates are considered low at <10%, and high at >30%, which means that English influence on Dutch remains overall low (4%), affecting mainly lexical loans. Contact between French and Middle English was prolonged, and started with the Norman Conquest in 1066. It took place over several stages, during which the contact varieties and intensity of contact evolved and changed. There was a peak of lexical borrowings around 1400, with lasting

consequences, since Present-day English still contains many French-origin words. French influx in Middle English was high (30%), and — at least according to more recent research (cf. mainly Rothwell and Ingham) — even left its trace in English syntax. This difference in borrowing rates and types of borrowings ties in with Thomason & Kaufman's (1991: 74–75; also see Thomason 2001: 70–71) borrowing scale, according to which less intense contact settings are restricted to lexical borrowings, whereas more intense contact settings can additionally result in syntactic borrowings. The settings also differ in the reciprocity of contact, as explained in more detail in Sections 2.4.3.1. and 2.4.4.1. above: despite the fact that contact between English and Dutch is unlikely to be mutual (Doğruöz & Zenner 2013; Zenner & Van De Mieroop 2017), several researchers have claimed that Middle English has influenced Anglo French in turn (Schauwecker 2018; Schauwecker & Trips 2018). Also, the two contact settings differ in terms of what proportion of the population has (or had) active access to the source language, the contact situation in Medieval England having been described by Ingham (2020: 452) as "a bilingual speech community, at least among higher-status and/or educated individuals"[25]. This is vastly different in the Low Countries, where the population has so far stayed monolingual to weakly bilingual (Doğruöz & Zenner 2013, based on European Union 2006), and where contact with English is mainly indirect (Booij 2001). The degree of societal bilingualism is thus a crucial difference between the two contact settings. Another crucial difference concerns the attitude towards using other-language material in one's own language. Although people in the Low Countries generally have positive attitudes to English (Edwards & Fuchs 2018), some extent of language purism forms part of the language climate (cf. examples such as the 'Federation against Loan words'). Following Zenner, Speelman & Geeraerts (2012), such negative attitudes characterise weak language contact settings as found in Western Europe[26]. In addition, people in Flanders have a strong sense of perceiving the standard language as being the only virtually ideal way of speaking (Grondelaers, Delarue & De Sutter 2017: 330–332). This phenomenon is the result of a long history of promoting the standard language (cf. De Sutter 2017), and as a consequence many people — in Flanders specifically — have a strong awareness of how they speak, and how they should or should not speak. In Medieval England, in contrast, borrowing — and even multi-word switches to French and Latin — was a highly common and widely accepted phenomenon (Wright 1992, 1995; Rothwell 2000; Schendl 2000; Schendl 2004; Ingham 2011a; Schendl & Wright 2011), and it was indicative of linguistic knowledge. As seen in an example from Schendl (2000: 81), extensive code-switching was even attested in

---

[25] However, as explained in Section 2.4.3.2. this is not a generally accepted claim.
[26] Purist attitudes are also found in minority languages: language users of Swedish in Finland, for instance, are considerably more hostile towards foreign influence than language users of Swedish in Sweden, where Swedish is a majority language (Graedler & Kvaran 2010).

letters to the king, which reveals the extent of its acceptance in society. Additionally, there was a less strong need for a unified dialect in Medieval England (Brinton & Arnovick 2011) than in the Low Countries. Another difference between the two contact settings is which part of the population speaks or spoke Present-day English and Anglo French. Whereas "it will be difficult to find a monolingual Dutch speaker under the age of 50 in the larger cities" in the Netherlands (Thomason 2001: 31), the lower classes in Medieval England remained largely monolingual (Schendl 2000). Another factor which distinguishes the contact settings under investigation is the differences between the morphology in source and recipient language. That is, Present-day English has less rich verbal and adjectival morphology than Present-day Dutch (cf. Chapters 3. and 4.), which means that inflections need to be added to (rather than left out from) English loans. The opposite tendency can be seen in the French-Middle English contact setting, since French generally has richer morphology than Middle English, which means that some inflections can be left out from French loans. The last difference does not concern the contact settings, but rather the type of data used in the present analyses. While the analysis for Dutch is based on a spoken corpus, which may be more authentic, the analysis for Middle English is compelled to written texts.

In sum, whereas the languages in both contact settings are related and come from identical or related language families, the historical contexts are different in that the intensity and duration of contact, as well as the degree of societal bilingualism, differ vastly.

## 2.5. General aims and hypotheses

The present thesis is concerned with loan word accommodation in two different contact settings, and — as indicated in Chapter 1. — the aim is two-fold.

(i)     In Part II, we will prove that — although loan words in direct insertion are treated like native words (Wohlgemuth 2009) — direct insertion may be subject to probabilistic biases. A pilot study on English loan verbs in Present-day Dutch had already revealed that the use of loan verbs is probabilistically biased to certain categories compared to native verbs, such that loan verbs are favoured in non-finite and uninflected forms (De Smet 2014). To this end, we will analyse loan word accommodation in the English-Dutch and French-Middle English contact settings. Both contact situations will demonstrate the existence (and persistence) of 'loan word accommodation biases', more specifically in verbs (Chapter 3.) and adjectives (Chapter 4.). Chapters 3. and 4. will shed more light on the nature of accommodation biases by showing that syntactic biases are consistently

stronger than morphological biases. Chapter 5. will discuss the potential causes of such biases: the main cause discussed in this thesis is that biases can be used as facilitative strategies to lower the increased processing cost associated with loan words undergoing morphosyntactic integration. By showing that loan words — even under direct insertion — are biased towards specific usage categories in their recipient language, this thesis will shed light on the nature of lexical borrowing as well as reveal constraints on loan word accommodation which have not been recognised so far. Notice that the focus of this thesis is on the integration of loan words in the language system (i.e. conventionalisation) rather than on individual differences between language users (i.e. entrenchment).

(ii)     In Part III, we will examine the potential *indirect* effects of language contact with French on (internal developments in) the syntax of English. We adopt the term 'indirect' (cf. Section 2.4.1.1.), since researchers have traditionally claimed that contact with French brought about no or little transfer of syntactic structures into Middle English. However, Middle English incorporated a substantial number of French loan words. Therefore, it is plausible that accommodation biases associated with French loans interacted with and impacted ongoing internal developments in the English language (cf. Hockett 1969: 414; Thomason 2001: 62). Diessel (2007: 117), for example, has suggested that "[s]mall biases in language production can lead to diachronic change". This is especially plausible for Late Middle English (1350–1500), comprising both the period of most intense French influence and its immediate aftermath (i.e. Early Modern English). Since French loan verbs in Middle English are favoured in non-finite forms compared to native English verbs (non-finite bias, cf. Part II), it is expected that the strong rise of non-finite constructions in the Late Middle English period is partially due to non-finite forms being exploited to accommodate French loan verbs. More specifically, the effect is tested on two verbal periphrastic constructions which came to a steep rise during Late Middle English and its aftermath, namely *do*-support (Chapter 6.) and light verb constructions (Chapter 7.). If the findings are indeed borne out, this will refine our current understanding (i) of a neglected aspect of contact-induced language change, namely changes which involve no direct transfer but which instead arise as a result of transfer and are thus indirect, and (ii) of some major trends in the history of English which caused the syntactic structure of English to diverge from that of the continental West-Germanic languages (e.g. McWhorter 2002), for instance in its high productivity of non-finite verb forms.

# Part II

# Existence, persistence and causes of accommodation biases

# Chapter 3

# Accommodation biases in verbs[27]

## 3.1. Introduction

As became clear from Section 2.3., loan words entering a recipient language typically accommodate to the grammatical template of that language (e.g. Poplack, Sankoff & Miller 1988; Muysken 2000; Wichmann & Wohlgemuth 2008). Four main loan word accommodation strategies have been identified by Wichmann & Wohlgemuth (2008) and Wohlgemuth (2009) in their typological research on verbs. The most common strategy cross-linguistically is direct insertion, where recipient-language inflections can be added directly onto the loan stem. This means that morphosyntactic integration is immediate and thus differs from phonetic integration, which is gradual (Poplack, Sankoff & Miller 1988). Due to the high frequency of direct insertion as well as the low integrational effort linked to it, Wohlgemuth (2009: 291) has concluded that the borrowing of loan verbs is not constrained by inflection. In this chapter, we verify this assertion by looking specifically into the accommodation of loan *verbs*, which is a complex phenomenon (cf. Section 2.3.1.). Most loan verbs undergo the following (optional) five processes when entering their recipient language (Wohlgemuth 2009: 56):

- assigning a loan verb to the word class 'verb'
- assigning to an inflectional class
- assigning to a classifying verb or an inflecting verb (in complex predicates[28])
- assigning valency
- attaching inflectional morphology

Regarding the borrowability of verbs, hierarchies of borrowability generally indicate loan verbs as being less borrowable than nouns (e.g. Muysken 1981; Poplack, Sankoff & Miller 1988; Matras 2007, 2009; Winford 2010; Hock & Joseph 2019), because verbs are said to be conceptually more complex than nouns (Meillet 1921;

---

[27] This chapter is based on a paper in the journal *Transactions of the Philological Society* (Shaw & De Smet 2022), and on parts of De Smet & Shaw (subm.). A pilot study of the Dutch data has been published in De Smet (2014).

[28] Complex predicates will been defined in Section 7.2.1.

Matras 2007, 2009). For instance, verbs generally bear more inflections than nouns (Harris & Campbell 1995) as they are inflected for tense, mood, and person (Poplack, Sankoff & Miller 1988). More reasons have been provided in Section 2.2.2.1.

The focus of this case study will be on loan verbs in two particular contact settings: English loan verbs in Dutch and French loan verbs in Late Middle English (for more contextualisation and examples, see Sections 2.4.3. and 2.4.4.). In both cases, direct insertion is the dominantly used accommodation strategy (cf. Sections 2.4.3.3. and 2.4.4.3.). As explained in Section 1.1., a pilot study by De Smet (2014; also see Meerts 2013) has shown that English loan verbs in Present-day Dutch are significantly more frequent in compound verb forms (e.g. non-finite forms such as infinitives and past participles) than in non-compound verb forms (e.g. finite forms). Apart from the English-Dutch contact setting, the present study also includes an analysis for the French-Middle English contact setting, and for both contact settings it compares the distributional properties of loan verbs and native verbs over the usage categories under investigation. The analysis reveals that loan verbs under direct insertion can be used like their native equivalents, and that all verbal usage categories are open to potential loans. However, loan verbs do not enter with the same ease into all inflectional categories or — within an inflectional category — into all specific formal variants, which corroborates the findings by De Smet (2014). This observation shows that — even under direct insertion — loan verbs are sometimes more readily inserted in some usage categories than in others, which means that direct insertion itself is not entirely free from constraints. The probabilistic tendency of loan verbs to enter more readily into specific syntactic and inflectional categories will be referred to as *loan word accommodation biases* (or accommodation biases). Two types of accommodation biases in particular are identified: (i) loan verbs are more likely to enter into a non-finite than into a finite form (*non-finite bias*[29]) and (ii) loan verbs are more likely to enter into a morphologically unmarked than into a marked form (*markedness bias*). This is a crucial correction to Wohlgemuth's (2009: 291) argument that, since direct insertion is the most frequently used accommodation strategy, inflection is not an obstacle to verbal borrowing. Instead, "apparent difficulties with loan verb accommodation are probably rather due to other, extralinguistic factors" (Wohlgemuth 2009: 291). In this manner, our findings add to existing evidence coming from — among others — Harris & Campbell (1995: 135), who state that nouns are more borrowable than verbs because they "tend to have fewer morphosyntactic markings than verbs, making loans easier to assimilate in this category". This implies that the presence of verbal inflections complicates borrowing. Typological research by Schultze-Berndt (2017) has previously illustrated this for Kriol, an English-lexified Creole language and *lingua franca*, and Jaminjung, a Northern Australian language.

---

[29] The concept of 'non-finite bias' is synonymous to 'finiteness bias', as used in Shaw & De Smet (2022).

Loan verbs from Kriol entering Jaminjung are without exception "integrated in the class of Uninflecting Verbs [UVs] rather than the class of Inflecting Verbs [IVs]" (Schultze-Berndt 2017: 265). She has added that the category of UVs is characterised by greater ease of borrowing (both into and from the category) than the category of IVs (Schultze-Berndt 2017: 265), yet she has not provided examples of this phenomenon, nor has she expanded on it. Since loan verbs are more prevalent in some usage categories than in others, it is possible that language users try to reduce the processing cost which they pay when integrating loan words in their recipient language. This argument will be further developed in Sections 3.4. and 5.3.4.

We will now first present the case study on the *Corpus Gesproken Nederlands* data for the English-Dutch contact setting (Section 3.2.). The French-Late Middle English contact setting, based on data from the PPCME2 and the *Helsinki Corpus*, will be discussed in Section 3.3. Both cases first present the methodology followed by the findings, which will substantiate our argument on loan word accommodation biases. In Section 3.4. we will frame the findings and initiate a potential theoretical explanation as to why loan word accommodation biases occur in loan verbs in the two contact settings under investigation.

## 3.2. English-Dutch contact setting

### 3.2.1. Data and methodology

#### 3.2.1.1. Data extraction

This case study follows the methodology developed in the pilot study by De Smet (2014). We first generated a list of lexical English verbs based on the part-of-speech-tagged *British National Corpus*, abbreviated BNC (Bodleian Libraries 2007). This was done by using a Perl script, which identified all verb tokens in the BNC. Based on that list, we then formed potential Dutch infinitives by adding the infinitival -(e)n-marker to the English stems, as in *reply-en*. The consonant immediately preceding -(e)n was doubled whenever required according to Dutch spelling rules, i.e. generally whenever the consonant follows a short single vowel, as in *bet-t-en*, derived from English *bet*. The hypothetical infinitives were then automatically searched for in the *Corpus Gesproken Nederlands* (Spoken Dutch Corpus), abbreviated CGN (Nederlandse Taalunie 2004). The CGN is a database and lemmatised corpus of Present-day Dutch as spoken in the Netherlands and Flanders. The corpus contains around 9 million words (5.6 million for Netherlandic Dutch and 3.3 million for Flemish), transcribed from 1,000 hours of adult speech recorded between 1998 and 2004. That the CGN contains spoken transcribed data makes the corpus of specific interest for this case study. First, the spoken character of the data neutralises the potential impact of Dutch spelling rules on the use of the loan verbs, which "can cause considerable problems in written forms" (Berteloot & Sijs 2002:

49). A particularly complex category are English verbs in *-e*, since the Dutch language does not have such verbs with base forms ending in *-e* (Berteloot & Sijs 2002: 49). A form such as *updatete*, for instance, orthographically seems to have four syllables, but is pronounced with three syllables only, as in /ˈʌpdeɪtə/. Due to its complexity, language users may avoid this type of forms in written texts[30]. Vendelin & Peperkamp (2006; also see Haugen 1950) have indeed shown that orthography influences loan word accommodation. Second, the spoken data in the CGN — and spoken data in general — is highly representative of authentic, spontaneous speech, where accommodation biases may be easiest to observe.

Searching for hypothetical Dutch infinitives in the entire CGN resulted in a list of 2,006 attestations containing 53 different relevant loan verbs, which are listed alphabetically in (39).

(39)    *blowen*[31], *boosten, callen, casten, catchen, checken, claimen, coveren, cutten, dealen, droppen, finishen, fixen, focussen, guessen, guiden, jumpen, keepen, liften* ('raise'[32]), *linken, managen, marketen, matchen, mixen, monitoren, moven, performen, plannen, poppen, pressen, printen, producen, promoten, pushen, quoten, relaxen, researchen, reviewen, runnen, saven, scoren, shaken, sharen, showen, smilen, sparren, splitten, starten, stretchen, switchen, testen, trainen, walken*

This data extraction method, using automatic extraction from the CGN, has some limitations. A first limitation is that the dataset at first included many false positives, including inexistent verbs (e.g. *ringen* falsely based on English *ring*), false friends (e.g. Dutch *loven* 'praise' is unrelated to English *love*) and cognates (e.g. Dutch *waken* 'be awake, guard' is not a loan but cognate to English *wake*), which were then filtered out manually. However, the manual clean-up was not a time-consuming process. A second limitation of this method is that it cannot distinguish between recent (e.g. *guessen*) and more established loans (e.g. *testen*), but it should be noticed that distinguishing between both is by definition an artificial process, and so is deciding on a cut-off point. A third limitation is that some loan verbs may have been overlooked by using this method, for instance since the verbal stems of the attestations are written differently in Dutch than in English (e.g. *fiksen* based on English *to fix*), or since they are based on English nouns instead of on verbs (e.g. pseudo loan verb *baseballen* based on English noun *baseball*). Another category which tends to be overlooked by using this method is the verbal category which is

---

[30] However, an advantage of using written data is that foreign material is sometimes used with quotation marks, which shows that the material is still associated with its source language (Andersen & Graedler 2020: 14). By working with a spoken corpus, we will not have access to this type of information.
[31] Note that *blowen* ('to smoke pot') may also be considered a pseudo loan verb, since English *to blow* is not used in the same sense as in Dutch.
[32] We only included attestations of *liften* in the sense of 'raise', as opposed to 'hitchhike'.

indirectly inserted into the language by adding an additional *-er*-suffix. This suffix is borrowed from French and is attested in verbs such as *format(t)-er-en*, *implement-er-en*, and *shock-er-en*, as seen in Berteloot & Sijs (2002: 48). However, we still opted for this data extraction method because of one major advantage which outweighs the limitations, namely that even the most recent loans — which have not been attested in dictionaries so far — can be retrieved. This means that the list presented in (39) is based on authentic spoken data, which is a major asset.

In order to investigate which tendencies are specific to loan verbs, the dataset containing English loans (n = 2,006) was matched to a dataset containing native Dutch verbs[33] (n = 1,813), which served as a control dataset throughout the analysis. Concretely, this means that each token of an English loan verb was randomly matched to a token of a native Dutch verb within the same frequency range (±25%). For example, native Dutch verb *knappen* ('to snap'), occurring 35 times in our dataset, was matched to English loan verb *printen* ('to print'), occurring 38 times. The total number of attestations in the dataset was 3,819.

The following high-frequency verbs were excluded from the Dutch verbs dataset, since — similar to the English verbs dataset — only lexical verbs were included: modals and auxiliaries *hebben* ('to have'), *mogen* ('may'), *moeten* ('must'), *kunnen* ('can'), *willen* ('want'), *worden* ('to become'), *zijn* ('to be') and *zullen* ('will'). We also eliminated separable verbs, such as *weggooien* ('to throw away'), since they may unnecessarily complicate data analysis: for instance, the finite categories of *weggooien*, as in *ik gooi weg* ('I throw away'), are categorised under the lemma of *gooien* ('to throw'), and not under *weggooien*. Matching the loan dataset to a baseline dataset was not only critical to examine which tendencies are typical of loan verbs, but also because of a historical tendency found in Dutch (and German) to replace simple past tense forms by perfect forms. Since this process (also called the *Präteritumschwund*[34]) — at least in German (Fischer 2018) — mainly affects low-frequency verbs, and loan verbs tend to be low-frequency items, the matching process ensures that any effects relating to the *Präteritumschwund* can be interpreted correctly.

### 3.2.1.2. Data annotation

Data annotation was conducted automatically, based on the annotations for inflectional forms. A simplified overview of the Dutch verbal inflectional system, based on Dutch-origin regular verb *werken*, is presented in Table 1.

---

[33] Although this dataset does not contain any English loan verbs, it is possible that it contains some older, well-established loans from languages such as Latin and Greek.
[34] For more information on the increasing grammaticalization of the perfect, see Coussé (2013).

*Table 1: Overview of the verbal inflectional system in Dutch (regular verb* werken*).*

| Inflectional category | Person and number | Inflectional form |
|---|---|---|
| Present | 1sg. | *ik werk* |
| Present | 2sg. | *jij werk-t / werk jij* |
| Present | 3sg. | *hij/zij/het werk-t* |
| Present | Pl. | *wij/jullie/zij werk-en* |
| Past | Sg. | *ik/jij/hij/zij/het werk-te* |
| Past | Pl. | *wij/jullie/zij werk-te-n* |
| Infinitive | N/A | *werk-en* |
| Past participle | N/A | *ge-werk-t* |

It becomes clear from Table 1 that the verbal stem (*werk*) is used without inflections in the first (*ik werk* 'I work') and — in the case of inversion — second person present singular (*werk jij* 'do you work'). In the second and third person singular, suffix *-t* is added, as in *jij werk-t* ('you work') and *hij/zij/het werk-t* ('he/she/it works'). The present plural as well as the infinitive consist of stem and present plural marker *-en*, hence *(wij/jullie/zij) werk-en* ('(we/you/they) work'). In the past, *-te* marker is added in the singular (*ik werk-te* 'I worked'; *jij werk-te* 'you worked'; *hij/zij/het werk-te* 'he/she/it worked'), and in the plural an extra plural marker *-n* is required (*wij werk-te-n* 'we worked'; *jullie werk-te-n* 'you worked'; *zij werk-te-n* 'they worked'). In the past participle, circumfix *ge⟩...⟨t* is added onto the stem, hence *ge-werk-t* ('worked'). Although the circumfix consists of a prefixing and a suffixing element and thus contains two separate elements, it only marks a single category. It becomes clear from this table that Dutch generally has richer verbal morphology than English: Dutch has more different inflections in both the present and past tense.

Based on this inflectional system, we distinguished between the following seven usage categories (illustrated by means of English loan verb *checken*), which cover both grammatical class and inflections:

1. The **unmarked present singular**, as found in the 1st (e.g. *ik check*) and inverted 2nd person (e.g. *check jij*)
2. The **-t-marked present singular**, as found in the 3rd (e.g. *hij check-t*) and non-inverted 2nd person (e.g. *jij check-t*)
3. The *-en*-marked **present plural** (e.g. *wij/jullie/zij check-en*)
4. The **past singular**, marked by *-de/-te* (e.g. *ik/jij/hij, zij, het check-te*)

5. The **past plural**, doubly marked by past *-de/-te* and plural *-n* (e.g. *wij/jullie/zij check-te-n*)
6. The *-en*-marked **infinitive** (e.g. *check-en*)
7. The **past participle** marked by circumfix *ge*⟩...⟨d/*t* (e.g. *ge-check-t*)

Whereas categories 1 until 5 cover finite forms, which — in Dutch — are inflected for person, number and tense, categories 6 and 7 cover non-finite categories, which mainly occur in Dutch in the passive and perfect. The dataset does not include prenominal past participles, which are used adjectivally, as in example (40).

(40)    *de **gepensioneerde** vrouw*

'the retired woman' (CGN)

*Te*-forms (*Hij ligt te **slapen*** 'He is sleeping', lit. 'He lies to sleep') and nominalised infinitives were excluded from the dataset as well. An example of a nominalised infinitive is shown in (41), where infinitive *autorijden* 'to drive a car' is combined with *'t*, an abbreviation of definite article *het*.

(41)    ging goed hé 't uh ***autorijden**?*

'It went well, um, driving, didn't it?' (CGN)

The present participle, a non-finite verb form, was not included in this study either due to its overall low occurrence rates in the Dutch language. An example of a present participle (*wervelend* 'whirling') is provided in (42).

(42)    *de **wervelende** sneeuw beperkte het zicht tot minder dan tien meter*

'The whirling snow reduced visibility to less than ten metres.' (CGN)

### 3.2.1.3. Data analysis

The paired sample for English loan verbs and Dutch native verbs was visualised in the shape of a mosaic plot using the R-packages "ggmosaic" (Jeppson et al. 2021) and "ggplot2" (Wickham 2016). This allows for a comparison of the rates of loan versus native verbs in the seven usage categories described in 3.2.1.2. Section 3.2.2. will explain in detail how to interpret such mosaic plots. We also calculated p-values for each usage category individually, using a Fisher's exact test (Levshina 2015: 214). This statistical significance test assesses whether there exist non-random associations between variables: is there enough evidence to show that a difference between the proportions of categories in two groups is not equally distributed? The result is expressed by means of a p-value, which indicates how likely it is that a given effect arises due to chance alone. The lower the p-value, the smaller the chance that the effect can be ascribed to chance, hence the higher the significance. In the p-value, two aspects are taken into account: to what extent the proportions are different, and how

much data there is. This means that the outcome of the Fisher's exact test should be considered with those two aspects in mind, and that a lack of significance can either point to the differences in proportions between categories not being large enough, or to the sample size being too small. Differences in proportions are traditionally considered significant at $p < 0.05$, and this significance threshold will be applied throughout this thesis as well. Fisher's exact tests are typically used for smaller sample sizes than Chi-square tests. In the present analysis, the Fisher's exact test compared the target category (e.g. present singular) to the remainder of the aggregate dataset (e.g. all categories except the present singular).

## 3.2.2. Findings

The data for this case study were visualised by means of a mosaic plot, which is defined by Nuzzo (2021: 531) as "a simple tool for visualizing, exploring, and displaying categorical data". It is a type of bar chart which can be used for visualising cross-tabulated data (Levshina 2015: 199), and which came to existence as a "graphical analogue of multivariate contingency tables" (Hofmann 2000: 23). The term 'mosaic plot' refers to "each cell of a corresponding table [being] visualised by a tile" (Hofmann 2000: 23), hence resembling a mosaic. Although mosaic plots are still not well-known among most researchers, they have the benefit of showing the relationship between two (or more) categorical variables at the same time, categorical data being "categories or labels rather than numerical measurements" (Nuzzo 2021: 531). Moreover, mosaic plots enable researchers to visualise rich data in a limited amount of space (Nuzzo 2021: 532), for instance by providing information on the frequency distribution of the data.

In this specific case, the mosaic plot visualises the distribution of English loan verbs over the different usage categories as compared to the control set of Dutch verbs. The seven usage categories with their abbreviations as printed in the mosaic plot are listed here below:

1. The unmarked simple present singular (abbreviated as "Pres.sg.")
2. The -*t*-marked simple present singular (abbreviated as "Pres.sg.-t")
3. The -*en*-marked simple present plural (abbreviated as "Pres.pl.")
4. The simple past singular (abbreviated as "Past.sg.")
5. The simple past plural (abbreviated as "Past.pl.")
6. The -*en*-marked infinitive (abbreviated as "Inf.")
7. The past participle (abbreviated as "Past.part.")

The mosaic plot in Figure 2 provides information about (i) the frequency distribution of the loan and control verbs in the different usage categories (vertical) and about (ii) the number of observations for each usage category, as can be observed from the width of the bars (horizontal). Thus, the more observations, the larger the cell. For

example, the past plural in Figure 2 is more frequent in control verbs than in English-origin verbs, and is the least frequent usage category in the overall dataset. The vertical dashed line in the middle of the graph is set at 50% and represents a perfectly 'bias-free' distribution of loan and control verbs, which have been frequency-matched in the data extraction procedure (cf. Section 3.2.1.1.). If the dataset is free of accommodation biases, the distribution for verbs of both origins will roughly coincide with the vertical line; however, if the dataset is not free of biases, the distribution for both origins will diverge accordingly.
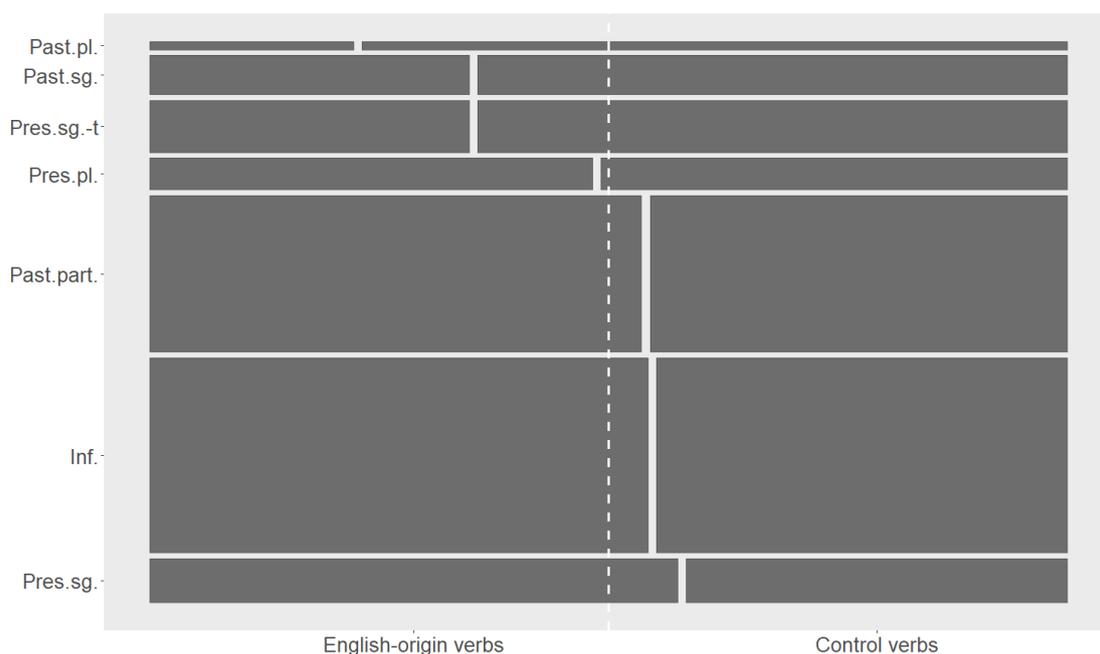


*Figure 2: Distribution of usage categories for English-origin verbs and control verbs (n = 3,819) with vertical dashed line set at 50%.*

As becomes clear from Figure 2, the distribution of English-origin verbs and control verbs does not coincide with the vertical dashed line. The loan verbs are, therefore, not entirely free from accommodation biases. The vertical divisions for each usage category reveal that English-origin verbs are overrepresented among some usage categories, whereas they are underrepresented among others. Starting from the bottom of the plot, loan verbs are disproportionally more frequent in the present singular, infinitive, and the past participle[35] as compared to the control verb dataset. This becomes clear from the width of the bars for the three categories. The effects

---

[35] Dutch has a regular rule of final devoicing in past participles, which may affect the choice of the simple past versus the past participle. However, since past participles are actually being favoured in loan verbs we suggest that an expansion on this topic is not needed.

for the usage categories are strongly significant, with p = 0.006 for the unmarked present singular, p < 0.001 for the infinitive, and p = 0.005 for the past participle. Moving to the top of the graph, the width of the bars reveals that English-origin verbs are disproportionally less frequent than the control group in the other four categories: the present plural, the -*t*-marked present singular, and the past singular and plural. The effect for the present plural (p = 0.58) is not significant, which may be because either the proportions of present plurals in loan verbs and native verbs are not different enough, or because there is not enough data. It may also be a combination of both reasons. The other three effects are all significant, with p < 0.001 for the -*t*-marked present singular, p < 0.001 for the past singular and p = 0.000 for the past plural. That present plural forms are quite common in French loan verbs may be due to the formal identicalness of present plural forms to infinitival forms, both ending in -*en*. Previous research has indeed shown that newer forms appear more often if they formally resemble successful existing (or other newer) forms (Naro 1981; De Smet 2012).

Based on the findings for Figure 2, two main observations can be drawn. First, inflection seems to be disfavoured in English loan verbs entering Dutch. This can be seen in the unmarked singular present, which consists of the verbal stem without inflection (e.g. *check*), being the most frequent inflectional category used by loan verbs. This is in stark contrast with the past plural, which is morphologically more complex since it receives both past and plural marking (e.g. *check-te-n*), and which is the least frequent inflectional category used in English loan verbs. This confirms the tendency of language users avoiding the addition of inflections on loan words, as found in the literature (e.g. Harris & Campbell 1995; Sijs 2005; Schultze-Berndt 2017: 265). For instance, an English loan verb would be more common in example (43), where the verb is used without inflection, than in example (44), where the verb needs to be inflected for the past plural.

(43)　*'t enigste nadeel is soms uh als er zo organisaties zijn van uh theaterinstellingen of wat dan ook die rekenen daar dan op dat ik regelmatig mijn e-mail* **checkØ**.

'The only disadvantage is that sometimes, um, when there are like organisations of, um, theatre companies or, um, whatever, they count on me checking my emails regularly (lit. they count on it that I regularly check my e-mail).' (CGN)

(44)　*ze* **scoorden** *ook goed bijvoorbeeld bij de Test Aankoop-uh-enquête u in verband met dienstverlening*

'They also scored well for instance on the Test Aankoop-um-survey, um, in terms of customer service.' (CGN)

In what follows, this tendency for loan verbs to occur in inflectionally unmarked contexts will be consistently referred to as the *markedness bias*.

A second observation which is suggested by Figure 2 is that English loan verbs in Dutch are generally favoured in non-finite verb forms, namely the infinitive and the past participle, which are typically used in compound forms (i.e. with an auxiliary). Admittedly, the unmarked present singular, where English loans are most frequent, is a finite verb form, but its high frequency could be explained by its formal resemblance to the originally borrowed form (*check*) and its absence of inflection, as described above. The preference to use loan verbs in non-finite forms may be surprising at first, since non-finite forms can be complex from a morphosyntactic point of view (e.g. past participle *ge-check-t*). However, they can be linked to the light verb strategy (Wohlgemuth 2009), one of the loan verb accommodation strategies described in Section 2.3.1.3. This strategy requires relatively low integrational effort, since it combines a non-finite form with a light verb, carrying the grammatical information, such as *he **took** a walk*, where *took* carries the information on person (third person), number (singular), and tense (past simple). Although the auxiliaries in the dataset — accompanying the non-finite forms — are not identical to light verbs, they operate similarly in carrying the grammatical information of the verbal clause. In *he **was** walking*, for example, it is auxiliary *was*, and not lexical *walking*, which provides information on person (third person), number (singular), and tense (past simple). In De Smet & Shaw (subm.; also see De Smet 2014) have, therefore, argued that non-finite forms generally carry less functional weight than finite forms since the lexical and grammatical components of the verb are more spread out. That non-finite forms carry less functional weight than finite forms is also seen in first-language acquisition research, where children typically use verbs in compound forms, such as infinitives and past participles, before they acquire the finite use of the verb (e.g. Wexler 1994; Kampen & Wijnen 2000). This tendency will be elaborated on in Section 5.3.1. In our dataset, an English loan verb would hence be more common in an example such as (45), a non-finite form, than in an example such as (46), a finite form.

(45)     *inderdaad Vanderstraeten die hier als uh zevende <u>zal</u> **finishen**.*

         'Indeed Vanderstraeten who will finish here, um, seventh.' (CGN)

(46)     *was wel vaag ja die spacecake [...] 't ging toch wel harder dan uh als je gewoon **blowt** of zo.*

         'It was quite blurry yeah that space cake [...] it went much faster than, um, if you had just smoked pot or something.' (CGN)

The tendency for loan verbs to occur in non-finite forms will be referred to as the *non-finite bias* throughout the rest of this thesis.

In this section we have found that, although English loan verbs can generally be used just like native Dutch verbs, their behaviour diverges from that of their native equivalents in that they are more easily adopted in specific usage categories (cf. De Smet 2014). We can conclude that English loan verbs in Dutch are subject to *loan word accommodation biases*. The two types of accommodation biases observed above — the markedness and the non-finite biases — raise the question whether such biases can only be found in the English-Dutch contact setting, or whether they also occur in other contact settings. The next section will look at the French-Middle English contact setting, which is in some ways similar to the English-Dutch contact setting. The focus will be specifically on the Late Middle English period.

## 3.3. French-Late Middle English contact setting

### 3.3.1. Data and methodology

#### 3.3.1.1. Data extraction

The analyses for this case study on Middle English drew on two different data sources, which are discussed separately in Sections 3.3.1.1.1. and 3.3.1.1.2. The first and main data source is the *Penn-Helsinki Parsed Corpus of Middle English (version 2)*, abbreviated PPCME2 (Kroch & Taylor 2000), which is a sub-corpus of the *Penn Parsed Corpora of Historical English* (Kroch 2020). It contains prose texts from a wide variety of genres, such as fictional texts, travelogues, homilies, and handbooks, in order to minimise the effects of specific discourse traditions on the use of syntactic forms (cf. Kabatek, Obrist & Vincis 2010: 1). The text samples in the PPCME2 are based on the samples included in the *Helsinki Corpus of English Texts* (Rissanen et al. 1991), but the former are considerably larger. In particular, two texts (i.e. *The Parson's Tale* and *Mandeville's Travels*) were retrieved in order to conduct a full-text analysis and gain more insight into both the markedness and non-finite biases. The second source of data is the *Helsinki Corpus of English Texts*, abbreviated HC (Rissanen et al. 1991), a diachronic corpus compiled between 1984 and 1991. The text samples from the HC were used in a follow-up study building on a large body of texts, which exclusively focused on the non-finite bias.

*3.3.1.1.1. The Parson's Tale and Mandeville's Travels*
The main data for the analysis were the PPCME2 versions of *The Parson's Tale* and *Mandeville's Travels*, two late 14th-century texts written in prose. The first text, *The Parson's Tale*, was originally written by Geoffrey Chaucer, but is based on Latin sources (Kroch & Taylor 2000). It is a 30,626-words religious treatise written in the iambic pentameter (Encyclopædia Britannica n.d.). Written around 1390, it forms part of Chaucer's famous *Canterbury Tales*. The author of the second text, *Mandeville's Travels*, has never been identified, but it is known that the text has been rendered in approximately 1400. *Mandeville's Travels* is a 51,715-words fictional travelogue,

translated from an immensely popular French text by Jehan de Bourgogne (1356/7) and based on two authentic travelogues: one by Guiollaume de Boldensele (1336) and the other by Odoric de Pordenone (1330) (Kroch & Taylor 2000). As such, *Mandeville's Travels* is a compilation of several travelogues (Van Tricht 2010: 17). Although the author, who remains anonymous, "writes very good English, [he] often misunderstands the French text" (Kroch & Taylor 2000). The text has three functions: it has a geographical interest, serves a didactic principle, and is entertaining (Van Tricht 2010: 18). It consists of two relatively independent parts: in part one, the author describes his fictional travels to Jerusalem, through Hungary, the Balkan peninsula, and Constantinople, and in part two he describes his travels through the "Far East" (Van Tricht 2010: 12). According to Van Tricht (2010: 17), researchers do not generally agree on the subgenre of *Mandeville's Travels*, which has been classified as a travel novel, a guidebook, etc.

*The Parson's Tale* and *Mandeville's Travels* are highly comparable texts: not only were they written in the same dialect, namely the East Midlands dialect, they were also written around the same time, namely when the rates of French loan words in Middle English surged drastically (Jespersen 1905; Dekeyser 1986). It is not coincidental, then, that both texts have a high incidence of loan words from French. Both texts were also written by authors who were fluent in French, Mandeville having translated a travelogue from French to Middle English and Chaucer having been "as fluent in French as in the Middle English of his time" (Lumiansky 2019: para. 4). According to Jespersen (1905: 92), Chaucer — like some other writers of the South — even used "a far greater number of French words than most other writers of this time". As a consequence of those similarities, the inflectional systems used in both texts are comparable. Inflections in Middle English were less regularised than in English today, since Middle English spelling had not been standardised yet[36]. As a result, the inflectional system is text-dependent. By means of illustration, Table 2 represents the text-specific idealised inflectional system as found for the regular finite forms in *The Parson's Tale* and *Mandeville's Travels*. Note that orthography and phonetics were closely related.

---

[36] English orthography only started to stabilise after William Caxton printed his first book (*Le Recueil des Histoires de Troyes*) in 1471 (Duff 2017). We are also well aware that none of the punctuation markers in our dataset are original.

*Table 2: Idealised maximal inflectional system for East Midlands finite forms in*
Mandeville's Travels *and* The Parson's Tale.

|  | Present | | | | Past | |
|---|---|---|---|---|---|---|
|  | Ind. | | Subj. | | | |
|  | Sg. | Pl. | Sg. | Pl. | Sg. | Pl. |
| 1 | *-e* | *-en* | *-e* | *-en* | *-ed-e* | *-ed-en* |
| 2 | *-est* | *-en* | *-e* | *-en* | *-ed-est* | *-ed-en* |
| 3 | *-eth* | *-en* | *-e* | *-en* | *-ed-e* | *-ed-en* |

The system for the East Midlands dialect as shown in Table 2 — like the Late Middle English period in its entirety — is characterised by the deflection it was undergoing, especially in forms endings in -(*ė*)*n* (e.g. Pyles & Algeo 1982: 153–154; Rissanen 2000). A form which showed variation between marked and unmarked inflection is the plural of the simple present. The deflection process had probably already started before the Middle English period, was more advanced in Northern than in Southern England, and has been linked to Scandinavian influence (McWhorter 2002; Emonds & Faarlund 2014). Deflection in the East Midlands variety has been accounted for in Table 2, where all forms with potentially reducible inflectional endings (i.e. the ones in -*e* or -*en*) were shaded. Examples of verbal deflection are presented in (47) and (48), where *assemblen* occurs with and without final -*n* respectively.

(47)  *For the ravenes & the crowes & the choughes & oþer foules of the contree* **assemble<u>n</u>**
       *hem þere euery ʒeer ones*

       'For the ravens, crows, and choughs, and other fowls of that country, assemble there once every year.' (*Mandeville's Travels*, c1400, PPCME2)

(48)  *þan anon þei armen hem &* **assembleØ** *hem togydere*

       'And then later they arm and assemble together.' (*Mandeville's Travels*)

Deflection was also found in Middle English infinitives, which are formally identical to present plurals. An example of inflectional variation for the infinitive is provided in (49) and (50), where it is shown that both *maken* and *make* are attested in Late Middle English.

(49)  *if he wole* **make<u>n</u>** *a trewe and a profitable confessioun ther moste be iiij condiciouns.*

       'If [anyone] wants to make a true and effective confession, four conditions must be met.' (*The Parson's Tale*, c1390, PPCME2)

(50)  *Thou ne shalt nat eek* **makeØ** *no lesynges in thy confessioun for humylitee*

'Nor should you during confession tell any lies for the sake of humility.' (*The Parson's Tale*)

A complication for present plurals and infinitives is that many forms lose their final *-n*-suffix, but retain the *-e*-suffix, which was originally part of the *-en*-ending (cf. examples (48) and (50)). The interpretation of this form is ambiguous, since it could have served as an intermediate form in between the old /ən/-pronunciation and the fully reduced form with absence of all inflection. In that case, one may argue that the *-e*-inflection still served as an inflectional ending. However, this interpretation is not plausible, since the *-e*-inflection did not formally differ from other potential verbal inflections and thus no longer distinguished *-e(n)*-forms from other forms. Furthermore, sources do not agree on how this final *-e* may have been pronounced. We, therefore, suggest that present plural and infinitive forms ending in *-e* had already lost their distinctiveness, and we will put forms ending in *-e* on the same par with forms without *-e*.

Variation also existed in the singular of the present, but only in the first person, where Ø-marker (e.g. *walk*) can alternate with *-e*-marker (e.g. *walk-e*). The singular of the past — with the exception of the second person singular, which generally ends in *-ed-est* (e.g. *walk-ed-est*) — could be reduced as well. Examples of such variation between *-e* and zero for the simple past are provided in (51) and (52).

(51)  *Thilke* **suffrede** *Iesu Crist withouten grucchyng*

'Jesus Christ suffered those [wicked words] without grudging.' (*The Parson's Tale*)

(52)  *That* **suffredØ** *Crist ful paciently*

'Christ suffered that with absolute patience.' (*The Parson's Tale*)

The reduction of inflectional forms — but now for past plurals — is illustrated in examples (53) and (54). Whereas *turnen* ('to turn') in (53) is inflected for the past plural (*-yn*), in (54) it is not inflected.

(53)  *And they dedyn right so as she bad hem and* **turnedyn** *ayen*

'And they did right as she asked them and turned back.' (*Mandeville's Travels*)

(54)   *We wetyn neuere whedyr thei were left in the vale or they* **turnedØ** *ayen at the entre, but we herdyn neuere more of hem aftyr.*

'We never knew whether they were left in the valley or whether they turned back at the entrance, but we have never heard from them again.' (*Mandeville's Travels*)

As can be seen from the above examples, the deflection process taking place in Late Middle English allowed language users to reduce inflectional forms. In line with the markedness bias, it could thus be hypothesised that French loan verbs in Middle English were also more frequent in inflectionally reduced forms. Note that Anglo French generally has richer verbal morphology than Middle English, and Late Middle in particular.

Although we mentioned that the forms which could possibly be reduced are the simple present singular and plural, the infinitive, and the simple past singular and plural, the present case study excludes present singular forms from the analysis in Section 3.3.2.2. The reason is that inflectional endings in the present singular cannot always be reliably identified since the inflections come with two major complications. First, the function of the -*e*-suffix in the singular of the present can be ambiguous: although — in some cases — it serves as an inflectional ending marking the present singular, in other cases it serves as an orthographic feature indicating the quality of the vowel in the preceding stressed syllable. In example (55), for instance, *clepe* carries an -*e*-suffix. It could as well be an inflectional ending for the present singular as an orthographic feature indicating that the vowel in *clepe* is pronounced with a long /e:/ instead of the shorter /e/-vowel.

(55)   *And herfore it is þat I* **clepe** *þe miȝtes of a soule, som principal, & som secundary.*

'And it is for that reason that I call some powers of the soul principal and some secondary.' (*The Cloud of Unknowing*, a1425, HC)

The second reason why the interpretation of inflection in the present singular is problematic is that a number of French-origin verbs have an ever-present final -*e* attached onto their stems, as in *I deceive, I judge* and *I prove*. Schwa could, again, be an actual inflectional ending, but it could just as well be an indicator of the quality of the vowel in the stressed syllable (also see Caon 2002).

### 3.3.1.1.2. Helsinki Corpus of English Texts

To gain more insight into our findings on the non-finite bias, we also conducted a follow-up study with a larger and more diverse dataset. To this end, we extracted the 27 text samples included in the third sub-period (i.e. 1350–1420) of the HC. The corpus consists of approximately 450 samples written between 730 and 1710, thus covering the Old, Middle and Early Modern English periods, and is the oldest

diachronic corpus available for English (Rissanen et al. 1991). The third sub-period of the HC, called "M3", contains a wide variety of short text samples from different authors, dialects and genres (e.g. sermons, letters, and narrative fiction), with a total word count of 140,380 for prose texts. We did not include any poetical texts since metre and rhythm may influence word order and may, therefore, present a distorted picture of the use of Middle English finite and non-finite verb forms (Kemenade 1987; Fischer, De Smet & Wurff 2017). Additionally, inflection may be adapted to the rhyme scheme (e.g. Babcock 1914). Table 3 provides an overview of the 27 prose text samples analysed in this study, including their authors — if known — and word counts.

*Table 3: Alphabetic overview of the 27 prose text samples (including word counts) analysed in the follow-up study (total N = 140,380).*

|   | Text (Author) | Word count |
|---|---|---|
| 1 | *Aelred of Rievaulx's De Institutione Inclusarum* | 3,010 |
| 2 | *A Late Middle English Treatise on Horses* | 6,110 |
| 3 | *A Latin Technical Phlebotomy* | 3,580 |
| 4 | *Appeal(s)* (Thomas Usk) | 2,220 |
| 5 | *Boethius* (Geoffrey Chaucer) | 10,170 |
| 6 | *Confessio Amantis* (John Gower) | 5,230 |
| 7 | *Cursor Mundi* | 10,390 |
| 8 | *Handlyng Synne* (Robert Mannyng) | 8,050 |
| 9 | *Letter(s); London Letters* (Henry V) | 5,010 |
| 10 | *Mandeville's Travels* | 5,530 |
| 11 | *Petitions, Returns, Judgements, Testaments and Wills* | 11,650 |
| 12 | *Polychronicon Ranulphi Higden* (John Trevisa) | 5,950 |
| 13 | *Proclamations* | 1,520 |
| 14 | *The Benedictine Rule* | 2,380 |
| 15 | *The Brut or The Chronicles of England* | 8,160 |
| 16 | *The Canterbury Tales: The General Prologue,* | 9,090 |

|  | *The Wife of Bath's Prologue,* <br> *The Summoner's Tale,* <br> *The Merchant's Tale,* <br> *The Tale of Melibee* (Geoffrey Chaucer) | |
|----|------------------------------------------------------------------------------------------------|--------|
| 17 | *The Canterbury Tales: The Parson's Tale* (Geoffrey Chaucer) | 4,410 |
| 18 | *The Canterbury Tales: The Tale of Melibee* (Geoffrey Chaucer) | 2,460 |
| 19 | *The Cloud of Unknowing* | 15,690 |
| 20 | *The Equatorie of the Planets* (Geoffrey Chaucer?) | 6,200 |
| 21 | *The New Testament* (John Wycliffe & John Purvey) | 11,000 |
| 22 | *The Northern Homily Cycle (the Expanded Version)* | 7,280 |
| 23 | *The Old Testament (Genesis and Numbers)* (John Wycliffe) | 9,910 |
| 24 | *The Pricke of Conscience* | 5,910 |
| 25 | *The Prologue to the Bible* (John Purvey) | 3,020 |
| 26 | *Treatise on the Astrolabe* (Geoffrey Chaucer) | 6,840 |
| 27 | *Wycliffite Sermons* | 13,460 |

Table 3 requires three more side notes. First of all, this dataset slightly overlaps with the *Mandeville's Travels* and *The Parson's Tale* texts, since short samples of both texts are included in this dataset as well (cf. numbers 10 and 17). Second, the diplomatic texts listed in the table have all been written contemporaneously with their original manuscripts, i.e. between 1350 and 1420. However, the originals of *Aelred of Rievaulx's De Institutione Inclusarum*, *Handlyng Synne* and *Cursor Mundi* were written between 1250 and 1350. This means that — period-wise — they only just belong to the dataset, but since their manuscripts date back from the same period as the other texts, we did not exclude them. Third, the reason why samples of *The Tale of Melibee* occur in the dataset twice (cf. numbers 16 and 18) is that one sample is classified as fiction whereas the other one is a philosophical text. To conclude, we followed the corpus builders in their selection of M3 texts.

### 3.3.1.2. Data annotation

Both datasets were annotated manually due to the rich spelling variation which characterises Middle English. It should be noted that we — due to the constraints of this thesis — only annotated a 25% subset of the HC dataset; more concretely, we annotated each fourth line of text. The verbal tokens were first lemmatised, which

allowed us to determine token frequencies for all lemmas, normalised per thousand words. For *The Parson's Tale* and *Mandeville's Travels* this lemmatisation was calculated per text, whereas in the HC dataset it was calculated for the dataset as a whole. After having been lemmatised, the verbal tokens were then annotated for source language and usage category, and — the ones in *The Parson's Tale* and *Mandeville's Travels* — for inflectional form. When we encountered any comprehension-related problems, we consulted the original text in which the attestation occurred, often written in French or Latin. A general decision in the annotation process was to analyse verbal constructions from bottom to top: *he must be buried* was hence annotated as a past participle (*buried*) rather than a passive (*be buried*) or even bare infinitive (*be buried*).

To identify the source language of the attestations, we used the etymological information as attested in the digital version of the *Middle English Dictionary*, abbreviated MED (Lewis 1952–2001). This dictionary is part of the *Middle English Compendium* (McSparran 2000), contains approximately 3,000,000 quotations retrieved from primary sources, and covers the period of roughly 1175 until 1500. In those cases where the MED did not provide (sufficiently clear) information, we consulted the *Oxford English Dictionary*, abbreviated OED (Oxford University Press 2009), and the *Dictionnaire Historique de la langue française* (Rey 2016). The different source languages which we distinguished between were English, French and Romance. It is possible that the category of 'English' contains some verbs which have other origins, for instance Old Norse, but it can be guaranteed that this category mainly contains verbs of Germanic origin. Distinguishing between English and Scandinavian is not relevant for this analysis, since the focus is on French loan verbs, which were more recent in Middle English than the already more conventionalised Old Norse loan verbs. Examples of verbs categorised as English are *answeren* ('to answer'), *finden* ('to find'), *kepen* ('to keep'), and *seien* ('to say'). Tokens annotated as 'French' in our dataset were all attestations which could be reasonably linked to a French source (e.g. Anglo French, Central French, Norman French in MED), even if a Latin source is also possible. Since Latin influence was partially contemporary with French influence, distinguishing between French and Latin origin is often not possible (Onions, Friedrichsen & Burchfield 1966: viii). We, therefore, created a category called 'Romance' in which loans with both French and Latin origin were included. However, loan verbs of purely Latin origin were not included in the analysis. Instances of loan verbs categorised as French or Romance are *accorden* ('to agree'), *chesen* ('to choose'), *commanden* ('to command'), and *contenen* ('to contain'). In the analysis, Romance-origin verbs were included in the dataset of French-origin verbs. As to the usage category of the attestations, verbal tokens could either be categorised as finite or non-finite. The

categories which are finite are the simple present[37], simple past, and the imperative; inflections of regular finite forms have been provided in Table 2. The categories which are non-finite are the gerund, infinitive, and past and present participles. Middle English gerunds and present participles were formally identical and generally received an *-yng(e)*-suffix; however, this suffix was subject to spelling variation, as in *walk-yng*, *walk-ynge*, *walk-ing*, *walk-inge*. In cases subject to heavy French influence, the present participle sometimes had the French ending *-a(u)nt* or *-and* (Mustanoja 1960: 548). Infinitives received an *-en*-suffix (as in *walk-en*), and past participles an *-ed*-suffix (e.g. *walk-ed*). For inflectional form, the last annotation category, all attestations subject to the envelope of variation between inflection and zero-inflection were annotated.

Since only lexical verbs were included in the analysis, the following verbs were excluded from the dataset: auxiliaries *ben* 'be' and *haven* 'have', and the modals (i.e. *connen* 'can', *moten* 'must', *mouen* 'may', *shulen* 'shall', *willen* 'will'), which may have an incomplete set of non-finite verb forms and are thus not representative. We also excluded mixed-origin verbs consisting of elements of both English and French origin, of which *defoulen* ('to trample upon, pollute') is an example, since it combines a Romance prefix (*de-*) and a Germanic root (*foilen*). This type of verbs presumably results from semantic re-borrowing and is rare. Compounds, which are verbal attestations at a morphological level, were excluded as well, since we exclusively looked into the syntactic level. An example of such a compound is provided in (56).

(56)    *bedes*-**byddynge**
        prayers-praying.GER
        'praying of prayers' (*John of Trevisa's Polychronicon*, a1387, HC)

Unclear or ambiguous attestations, excerpts in French or Latin, and Present-day English meta-information were excluded as well. The same applied to loan translations (or calques), such as *for-saien* ('to renounce'), which is modelled on Latin *prae-dictum*. Last, we also excluded fixed expressions (referred to as 'idioms' in OED), namely *according to*, *that is to say*, and *quod*, where verb forms have been grammaticalized over time. After excluding the above verbal tokens, the dataset for *The Parson's Tale* and *Mandeville's Travels* contained 3,881 verbs, and the dataset for the HC contained 4,434 verbs. 22% (864 verbs) of the verbs in *The Parson's Tale* and *Mandeville's Travels* dataset were of French origin; for HC this was 18.7% (829 verbs).

---

[37] Since the distinction between the Late Middle English indicative and subjunctive can only be made in the second and third person singular (cf. Table 2), indicatives as well as subjunctives were included in this category.

### *3.3.1.3. Data analysis*

First data inspection was conducted by means of mosaic plots, using the R-packages "ggmosaic" (Jeppson et al. 2021) and "ggplot2" (Wickham 2016). We additionally used the Fisher's exact test — with significance set at $p < 0.05$ — to calculate the p-values of each individual usage category. The data were then analysed in the statistical software R. In order to predict the effect of source language on the use of finiteness and inflection, we carried out mixed-effects logistic regression models, using the R-package "lme4" (Bates et al. 2015). This type of statistical model was opted for since the dependent variable is binary and its outcome can be predicted by multiple independent variables. Frequency of French- and English-origin verbs was included in the regression models as well, since the data for this case study — as opposed to the one for Dutch — were not frequency-matched[38]. The findings section will report more information on how logistic regression models function and on how to interpret them.

## 3.3.2. Findings

This section examines whether French loan verbs in Late Middle English — like English loan verbs in Present-day Dutch — were subject to a non-finite (3.3.2.1.) and markedness bias (3.3.2.2.).

### *3.3.2.1. Non-finite bias[39]*

In this section we investigate whether French loan verbs in Late Middle English are disproportionally more frequent in non-finite than in finite categories. We have reasons to suggest that this may be the case, first of all since the languages in contact are similar to the language pair in the English-Dutch contact setting. Additionally, accommodation biases have not only been found for Dutch, but have also been hinted at elsewhere in the literature. Indeed, Fischer & Wurff (2006: 155) have argued that French loan verbs in Middle English may have been more frequent with *do*-support than native verbs. This tendency may have aided in avoiding inflecting loan verbs because *do*-support is a periphrastic construction allowing the lexical verb to be used as a non-finite form. However, since the rise of *do*-support constitutes the main focus of Chapter 6., the present chapter does not enter into more detail on this topic.

Preliminary data inspection of both *The Parson's Tale* and *Mandeville's Travels* and the HC datasets is conducted by means of mosaic plots, which represent the distribution of French-origin verbs versus control verbs in the seven usage categories under

---

[38] The reason why the methodology for the French-Middle English contact setting is different from the one for the English-Dutch contact setting is because of the different nature of both datasets. Whereas the Dutch data are easily searched automatically because spelling is consistent, they are impossible to search manually because loan verbs are relatively infrequent. For the Middle English data the reverse situation holds.

[39] Evidence of the non-finite bias in French loan verbs accommodating to Middle English has also been found in a follow-up study by Elter & Shaw (2022). In this study, the hypothesis is tested on a much larger body of texts than in the present thesis. This study will also be expanded on in Section 8.2.

investigation. The categories as well as their corresponding abbreviations (as printed in the mosaic plots) are repeated below:

1. The simple present (abbreviated as "Pres.")
2. The simple past (abbreviated as "Past")
3. The imperative (abbreviated as "Imp.")
4. The infinitive (abbreviated as "Inf.")
5. The present participle (abbreviated as "Pres.part.")
6. The past participle (abbreviated as "Past.part.")
7. The gerund (abbreviated as "Ger.")

The mosaic plots in Figure 3 and Figure 4 represent the distributional properties of verbs in *The Parsons's Tale* and *Mandeville's Travels* and in the HC respectively. Both graphs should be interpreted similarly to Figure 2: whereas the frequency distribution of loan and control verbs in the different usage categories is reflected vertically, the number of observations per usage category is reflected horizontally. However, a critical difference with Figure 2 is that verbs of French and English origin in this sample have not been frequency-matched. Additionally, the vertical dashed lines in Figure 3 and Figure 4 represent the overall share of French-origin verbs in the datasets, which have been set at 22% (*The Parsons's Tale* and *Mandeville's Travels*) and 18.7% (HC). A bias-free distribution in Figure 3 would hence coincide with the 22%-mark, and in Figure 4 with the 18.7%-mark.
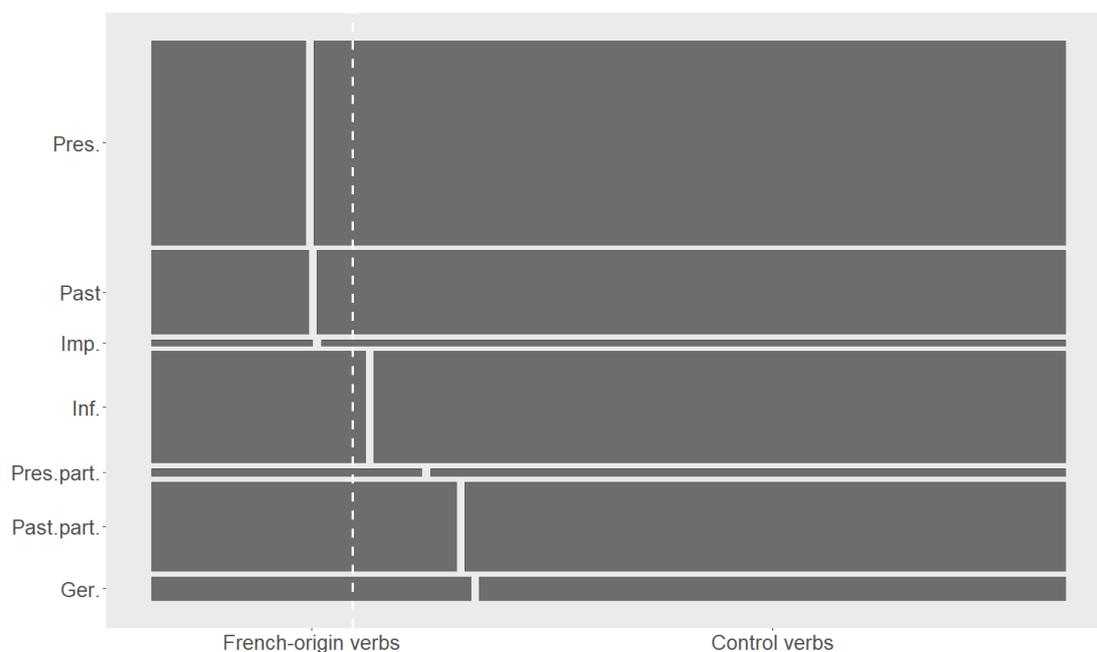


*Figure 3: Distribution of usage categories for French-origin verbs and control verbs (n = 3,881) in PPCME2 with vertical dashed line set at 22%.*
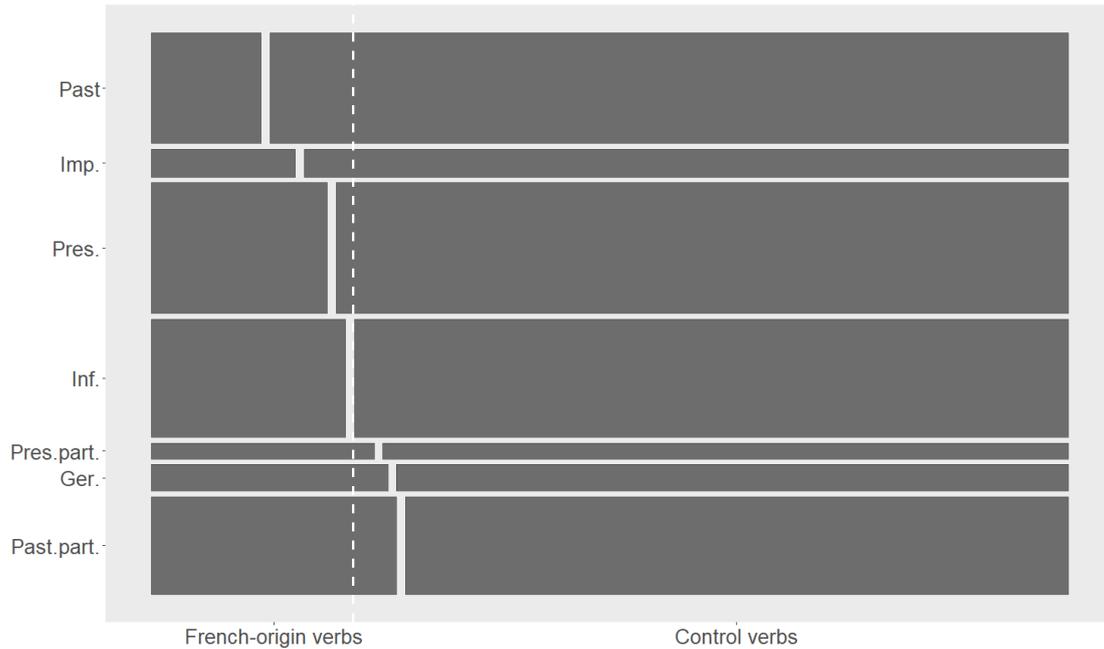
*Figure 4: Distribution of usage categories for French-origin verbs and control verbs (n = 4,434) in HC with vertical dashed line set at 18.7%.*

For Figure 3, first, French loan verbs are disproportionally more frequent in the gerund (p < 0.001), past participle (p < 0.001), present participle (p = 0.20), and the infinitive (p = 0.32). This can be seen from the vertical divisions for the French-origin verbs in the usage categories. However, French loan verbs are disproportionally less frequent in the imperative (p = 0.59), past (p = 0.001), and present (p < 0.001). Whereas the effects for the gerund, past participle, past, and present are significant, the effects for the infinitive, present participle, and imperative are not, but they go into the expected direction. The lack of significance may be explained by a lack of difference between the proportions across loan verbs and native verbs, or by a lack of data.

Similarly, French loan verbs in Figure 4 are overrepresented in non-finite categories and underrepresented in finite categories. However, the internal order among usage categories differs slightly. French-origin verbs in the HC dataset are uncommonly frequent in the past participle (p = 0.01), gerund (p = 0.003), present participle (p = 0.06), and the infinitive (p = 0.007), whereas they are uncommonly infrequent in the present (p = 0.47), imperative (p = 0.23), and the past (p < 0.001). Although the effects for the present participle, the present, and the imperative follow the expected trend, they are not significant. For the present the effect is likely too small to be significant, and for the present participle and the imperative the dataset likely does not contain a sufficiently large amount of data.

The two mosaic plots paint a highly comparable picture, albeit with slight differences among the exact frequency rates within the categories. In both datasets, French-origin verbs are overrepresented in non-finite categories, whereas they are underrepresented in finite categories. However, it should be noticed that the two figures above purely serve as a first indication of the findings, and not as a full-fledged analysis of the data. That is because Figure 3 and Figure 4 cannot account for possible frequency effects, the dataset not having been frequency-matched. Also, neither plot controls for the fact that our data were retrieved from different texts which may be subject to their own idiosyncratic features. To take those two complications into account, we additionally performed mixed-effects logistic regression analyses (cf. Levshina 2015: Ch. 12.) of the two datasets. A logistic regression model is a statistical technique used to predict "the relationships between a categorical response variable [also called dependent variable] with two or more possible values and one or more explanatory variables, or predictors [also called independent variables]" (Levshina 2015: 253). In the model for *The Parson's Tale* and *Mandeville's Travels* in Table 4, finiteness (non-finite: 0/finite: 1) serves as the dependent variable, whereas source language and lemma frequency serve as the independent variables. Concretely, we tested the effects of source language and lemma frequency on the use of finiteness. In order to control for potential author-, register-, or text-specific features hiding in *The Parson's Tale* and *Mandeville's Travels*, text was selected as a random effect[40]. Lemma as well was included as a random effect in order to account for possible lemma effects. By using random effects, which are described in Levshina (2015: Ch. 8.), we essentially took into account the values of text and lemma without looking at them as actual independent variables. We also ran an interaction (cf. Levshina 2015: 162–166) between source language and lemma frequency. This is critical to bear in mind, as the interaction can affect the main effects for source language and lemma frequency.

Table 4 presents the outcome of the logistic regression model, with the abbreviation 'Est.' standing for the coefficient estimates of the fixed effects. In cases where the coefficient estimate is positive, the correct interpretation is that the mean for the dependent variable will increase as the value for the independent variable increases; reversely, in cases where the estimate is negative, the mean for the dependent variable will decrease as the value for the independent variable increases. 'SE' in Table 4 is the abbreviation for the standard errors of the fixed effects, which represent the average distance of the observed values to the regression line. Hence, the lower the standard error, the more reliable the finding is. 'CI' (95%) stands for the confidence intervals of the coefficient estimates and helps to assess the significance of the findings. The output in Table 4 depicts both the lower and upper bounds of the confidence

---

[40] This method cannot account for potential intra-linguistic variation in authors' language production.

intervals. All numbers are rounded to two digits after comma. Note that lemma frequency is compared to the mean, and French-origin verbs to English-origin verbs.

*Table 4: Outcome of mixed-effects logistic regression model for finiteness in* Mandeville's Travels *and* The Parson's Tale *(n = 3,881).*

|  | Est. | SE | CI Lower bound | CI Upper bound |
|---|---|---|---|---|
| (Intercept)[41] | -0.05 | 0.13 | -0.49 | 0.38 |
| Source (French-origin) | -0.51 | 0.11 | -0.74 | -0.29 |
| Frequency | 0.40 | 0.04 | 0.33 | 0.48 |
| Source (French-origin): Frequency | 1.73 | 0.59 | 0.60 | 2.91 |

Table 4 corroborates that source language has a significant impact on the selection of finite versus non-finite forms, as suggested in Figure 3 and Figure 4: as the coefficient estimate (-0.51) for source language is negative, the dependent variable (finiteness) is less probable with verbs of French origin. This means that, as hypothesised, French loan verbs are significantly more likely to occur in a non-finite form, as in (57), whereas a native English verb is more likely to occur in a finite form, as in (58).

(57)  *if he had ben mair, I wot wel he wolde haue* **meigtened** *al hys ordinances*

'If he had been mayor, I know that he would have maintained all his assignments.' (*The Appeal of Thomas Usk against John Northampton*, 1384, HC)

(58)  *a-non after mete* **kom** *John Norhampton to John Mores hows, & thider* **kom** *Richard Norbury & William Essex*

'At noon after lunch John Northampton came to John Moore's house, and Richard Norbury and William Essex came as well.' (*The Appeal of Thomas Usk against John Northampton*, 1384, HC)

The positive coefficient estimate (0.40) for lemma frequency in Table 4 additionally shows that frequency has a significant effect on the use of finiteness in verbs: more concretely, low-frequency verbs are more frequent in non-finite forms as compared to high-frequency verbs. This is not surprising, since low-frequency and high-frequency verbs respond differently to variation and change. The model outcome also reveals that source language and lemma frequency interact, meaning that the two variables combined have a significantly larger effect on the use of

---

[41] The intercept represents the mean value for the response variable when all the independent variables are equal to zero. In this case, it represents the probability of the finite value in an English-origin verb when the independent variables are equal to zero.

inflection than the individual variables alone. More concretely, the preference to use French loan verbs in non-finite forms becomes even stronger in lower-frequency items than in higher-frequency items. A possible explanation is that French-origin lower-frequency verbs (e.g. *assenten* 'to assent', *consideren* 'to consider') are typically newer loan words, while French-origin higher-frequency verbs (e.g. *minishen* 'to reduce', *ordeinen* 'to put in order') are the better-established loan words (Langacker 1987; Bybee 2010). Lower-frequency items, and thus newer loan words, are harder to retrieve mentally than higher-frequency items since they have not reached the same point of entrenchment as higher-frequency items (Langacker 1987; Bybee 2010; Schmid 2016). Also, the processing cost of lower-frequency items is found to be higher than that of higher-frequency items (Berglund-Barraza et al. 2019; Desai, Choi & Henderson 2020). As such, language users may feel more of a need to resort to non-finite forms with lower-frequency items than with higher-frequency items, which will be discussed in Section 5.3.2. Moreover, that the bias is less pronounced in a higher-frequency, better-established loan than in a lower-frequency and more recent loan may suggest that accommodation biases weaken over time, but this topic will be expanded upon in Section 4.4.

In order to ascertain whether the non-finite bias also holds for a larger collection of texts, the same mixed-effects logistic regression model was applied to the HC data. We first ran a model with both source language and lemma frequency as main effects. When we additionally checked for an interaction effect between the two variables, the interaction turned out to be insignificant. Therefore, we ran a new model without the interaction, which is displayed in Table 5. The variables and random effects from Table 4 were maintained in this model.

*Table 5: Outcome of mixed-effects logistic regression model for finiteness in the 1350–1420 period of the HC dataset (n = 4,434).*

|  | Est. | SE | CI Lower bound | CI Upper bound |
|---|---|---|---|---|
| (Intercept) | -0.16 | 0.09 | -0.36 | 0.02 |
| Source (French-origin) | -0.21 | 0.08 | -0.37 | -0.05 |
| Frequency | 0.04 | 0.005 | 0.03 | 0.05 |

The negative coefficient value for verbs of French origin (-0.21) in Table 5 confirms the non-finite bias found for loan verbs as opposed to control verbs. However, the effects are less pronounced than in Table 4. Lemma frequency (0.04) has a significant impact on the choice between finite and non-finite verb forms as well, meaning that the non-finite bias is — again — even stronger in lower-frequency verbs. Different

from Table 4, source language and lemma frequency in Table 5 do not interact (cf. supra): therefore, the two variables combined do not have a significantly larger effect on the use of inflection than the individual variables alone. That the effects for the HC data are slightly less pronounced than the effects found for *The Parson's Tale* and *Mandeville's Travels* could pertain to the generally higher rates of French loans in *The Parson's Tale* and *Mandeville's Travels*. The HC data, in contrast, contains texts such as *The Old* and *New Testaments*, which have retained remarkably fewer traces of language contact with French.

### 3.3.2.2. Markedness bias

The present section examines whether the markedness bias found for English loan verbs in Present-day Dutch also applies to French loan verbs in Late Middle English. Since Middle English was undergoing deflection at the time (Section 3.3.1.1.1.), and inflectional marking on verbs was variable, the hypothesis is that loan verb status affects a verb's chances of receiving inflectional marking. We first look into deflection patterns in the present plural and infinitive (Section 3.3.2.2.1.), as well as the singular and plural of the simple past (Section 3.3.2.2.2.). Why the singular of the simple present was excluded from the analysis has been explained in Section 3.3.1.1.1.

*3.3.2.2.1. Simple present plural and infinitive*

In this first case study we look into the inflectional endings used in the present plural and infinitive, where morphologically unmarked forms, which carry zero- or -*e*-inflections, are contrasted to marked forms, which carry -*en*-inflections. The mosaic plots in Figure 5 illustrate this distribution.
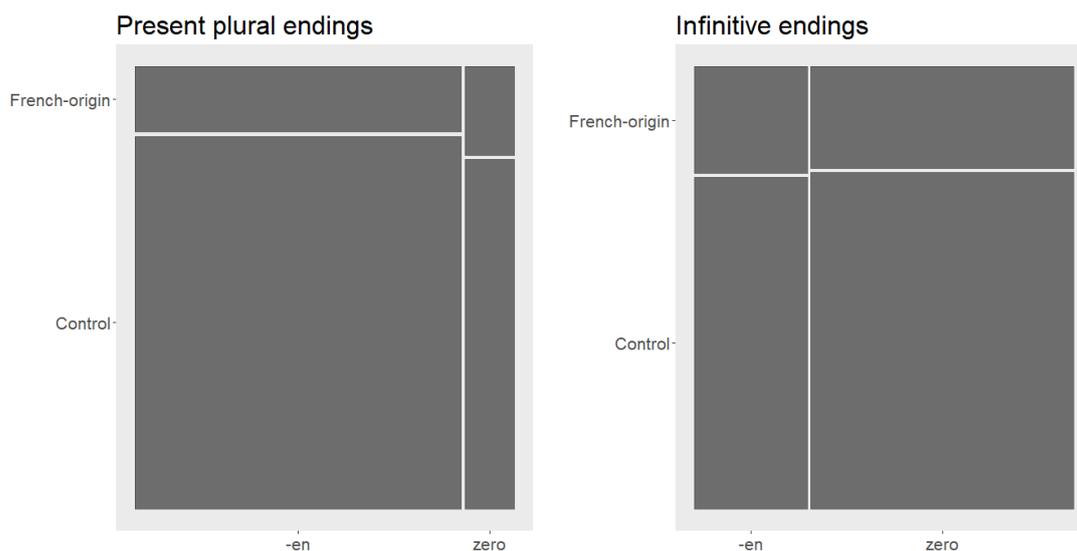


*Figure 5: Distribution of* -en- *and zero-endings for French-origin verbs and control verbs in present plurals (n = 718) and infinitives (n = 823) in* Mandeville's Travels *and* The Parson's Tale.

At first sight, French-origin verbs in the present plural appear to be slightly more frequent with zero-endings than with -*en*-endings, which can be inferred from the horizontal divisions for French-origin verbs. The reverse is true for infinitives, where French-origin verbs appear to be slightly more frequent with -*en*-endings than with zero-endings. Again, this can be inferred from the horizontal divisions. However, the differences in distributional properties between loan verbs and control verbs are minimal. Additionally, the vertical divisions in the mosaic plot reveal that patterns of deflection were much more common in infinitives than in present plurals, showing that infinitives were more innovative.

In order to gain more insight into the distributional properties of French-origin verbs versus control verbs we carried out a mixed-effects logistic regression model (for more information on this model, see Section 3.3.2.1.) for present plurals and infinitives separately. Inflectional variation (zero/-(*è*)*n*) serves as the dependent variable, whereas source language and lemma frequency serve as the independent variables. In the model of infinitives specifically, we included the type of infinitive — either a bare infinitive (e.g. *come*) or a *to*-infinitive (e.g. *to come*) — as an additional independent variable possibly affecting inflectional use. This is the main reason for creating two separate models instead of including the distinction between infinitives and present plurals as an independent variable in one model. As was done for Table 4 and Table 5, text was included as a random effect in order to take potential inflectional differences between *Mandeville's Travels* and *The Parson's Tale* into account. The outcome for present plurals and infinitives is presented in Table 6 and Table 7 respectively.

*Table 6: Outcome of mixed-effects logistic regression model for markedness in present plurals (n = 718) in* Mandeville's Travels *and* The Parson's Tale.

|  | Est. | SE | CI Lower bound | CI Upper bound |
|---|---|---|---|---|
| (Intercept) | 2.0 | 0.34 | 0.95 | 2.94 |
| Source (French-origin) | -0.39 | 0.34 | -1.05 | 0.29 |
| Frequency | -0.21 | 0.15 | -0.51 | 0.09 |

*Table 7: Outcome of mixed-effects logistic regression model for markedness in infinitives (n = 823) in* Mandeville's Travels *and* The Parson's Tale.

| | Est. | SE | CI Lower bound | CI Upper bound |
|---|---|---|---|---|
| (Intercept) | -0.97 | 0.24 | -1.69 | -0.23 |
| Source (French-origin) | 0.03 | 0.19 | -0.35 | 0.41 |
| Frequency | -0.05 | 0.10 | -0.26 | 0.14 |
| Type (*to*-infinitive) | 0.38 | 0.16 | 0.07 | 0.69 |

The model for infinitives (Table 7) reveals one significant effect: *to*-infinitives are more likely to be used with -(*e*)*n*-endings than bare infinitives, which is shown by the positive coefficient value (0.38). This means that *to*-infinitives are less progressive than bare infinitives. However, neither model reveals any significant effects for source language or lemma frequency affecting the deflection pattern of present plurals or infinitives. It is somewhat surprising that we do not even find frequency effects for the above tables, since researchers generally agree that phonetic changes, and in particular those involving phonetic reduction, are usually submitted to frequency effects (see e.g. Krug 2003; Bybee 2006).

Splitting up the two datasets in four, namely by looking at present plurals and infinitives in the two texts separately, we obtain the following four regression models: present plurals (Table 8) and infinitives (Table 9) in *The Parson's Tale*, and present plurals (Table 10) and infinitives (Table 11) in *Mandeville's Travels*. The variables and random effects from Table 6 and Table 7 were maintained here. The only difference is that the tables — instead of confidence intervals — now present the p-values as found in the model. The models for *The Parson's Tale* in Table 8 and Table 9 do not depict any significant effects. Whereas the type of infinitive was significant in the model for *The Parson's Tale* and *Mandeville's Travels* together, there is no effect in *The Parson's Tale* data alone.

*Table 8: Outcome of mixed-effects logistic regression model for markedness in present plurals (n = 156) in* The Parson's Tale.

| | Est. | SE | P-value |
|---|---|---|---|
| (Intercept) | 1.14 | 0.34 | 0.00 |
| Source (French-origin) | 0.50 | 0.51 | 0.33 |
| Frequency | -0.04 | 0.20 | 0.83 |

*Table 9: Outcome of mixed-effects logistic regression model for markedness in infinitives (n = 398) in* The Parson's Tale.

|  | Est. | SE | P-value |
|---|---|---|---|
| (Intercept) | -0.61 | 0.18 | 0.00 |
| Source (French-origin) | -0.17 | 0.27 | 0.53 |
| Frequency | 0.01 | 0.11 | 0.92 |
| Type (*to*-infinitive) | 0.22 | 0.22 | 0.31 |

The lack of significant effects may be due to the small sample sizes for *The Parson's Tale*. The model including only the *Mandeville's Travels* data does reveal a number of significant effects. For present plurals, source language (p = 0.03) has a significant effect on the use of inflection. More concretely, the negative coefficient for French-origin verbs (-0.94) shows that loan verbs are more likely to reduce inflectional endings than control verbs. The effect for frequency, however, is not significant. The output of this model is shown in Table 10.

*Table 10: Outcome of mixed-effects logistic regression model for markedness in present plurals (n = 562) in* Mandeville's Travels.

|  | Est. | SE | P-value |
|---|---|---|---|
| (Intercept) | 2.52 | 0.28 | <2e-16 |
| Source (French-origin) | -0.94 | 0.42 | 0.03 |
| Frequency | -0.30 | 0.23 | 0.20 |

For infinitives (Table 11), there is no significant effect for source language on inflectional form. However, the negative coefficient (-0.49) for frequency reveals a marginally significant frequency effect (p = 0.06) for higher-frequency verbs being favoured with zero-suffixes. The type of infinitive (p = 0.03) is again significant: the positive coefficient value (0.52) shows that *to*-infinitives are more frequent with -(*e*)*n*-endings as compared to bare infinitives.

*Table 11: Outcome of mixed-effects logistic regression model for markedness in infinitives (n = 425) in* Mandeville's Travels.

|  | Est. | SE | P-value |
|---|---|---|---|
| (Intercept) | -1.14 | 0.22 | 1.33e-07 |
| Source (French-origin) | 0.11 | 0.29 | 0.71 |
| Frequency | -0.49 | 0.26 | 0.06 |
| Type (*to*-infinitive) | 0.52 | 0.24 | 0.03 |

This section has provided some evidence of a markedness bias in French-origin verbs in Late Middle English, more particularly in present plural forms and infinitives. As shown in example (59), a French-origin infinitive would be more common in an uninflected form, such as *soupe*, than in an inflected form, such as *yeven*.

(59)    " *and he shal* **soupe** *with me* " *by the grete joye that I shal* **yeven** *hym* .

        "'And he shall dine with me" by the great joy that I shall give him.' (*The Parson's Tale*)

It should be noted that the case for present plurals and infinitives is not particularly strong: evidence was found exclusively for *Mandeville's Travels*, and the datasets used for this case study were small. The next section will present a second and more convincing case study.

### 3.3.2.2.2. *Simple past singular and plural*

Similar to present plurals and infinitives, past forms in Late Middle English were subject to deflection. Preliminary data inspection is shown in Figure 6, presenting the distribution of inflectional endings across French-origin verbs and control verbs[42] in both the past singular (zero versus -*e*) and the past plural (zero versus -(*e*)*n*).
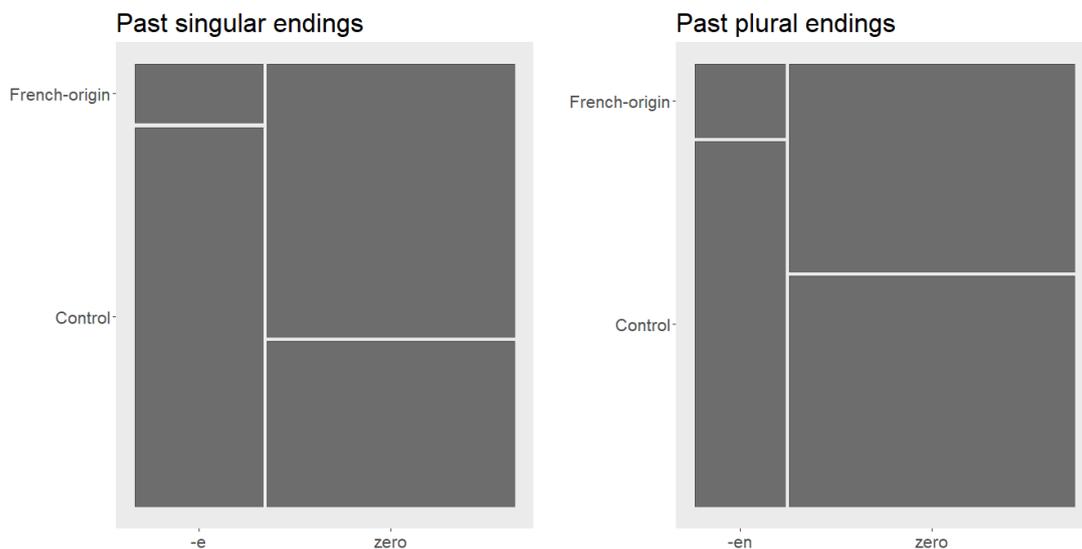


*Figure 6: Distribution of* -e- *and zero-endings for French-origin verbs and control verbs in the past singular (n = 197) and* -(e)n- *and zero-endings in the past plural (n = 25) in* Mandeville's Travels *and* The Parson's Tale.

---

[42] In this case, the set of control verbs only includes weak verbs, and no strong or irregular verbs. Since French loan verbs typically entered into Middle English as weak verbs, this decision ensures that French verbs can be compared to control verbs with the same past tense formation pattern.

It becomes clear from Figure 6 that French loan verbs in both the past singular and plural are more frequent without than with inflectional endings as compared to the weak control verbs. This can be inferred from the horizontal divisions. At first sight, the effects — and in particular the one for the past singular — seem pronounced. Zero-inflection is also generally more common in past singulars and plurals than inflection, which can be inferred from the vertical divisions. Whereas the dataset for past plurals only contains 25 attestations and is too small for any additional statistical testing, we performed a mixed-effects logistic regression analysis on the past singular data, with inflectional variation (zero/-*ĕ*) as the dependent variable, and source language and lemma frequency as the independent variables. Text was assigned as a random effect. The output of the model is displayed in Table 12.

*Table 12: Outcome of mixed-effects logistic regression model for markedness in*
*the past singular (n = 197) in* Mandeville's Travels *and* The Parson's Tale.

|  | Est. | SE | CI Lower bound | CI Upper bound |
|---|---|---|---|---|
| (Intercept) | -0.88 | 0.28 | -1.46 | 0.05 |
| Source (French-origin) | -1.50 | 0.44 | -2.41 | -0.67 |
| Frequency | 1.06 | 0.23 | 0.65 | 1.56 |

The negative coefficient value (-1.50) for source language in Table 12 corroborates the effect of source language on the use of inflectional endings in the past singular, as suggested in Figure 6: French loan verbs (e.g. *regnen* 'to reign') are less likely to be used with inflection than the weak control verbs (e.g. *haven* 'to have'). This phenomenon is illustrated in examples (60) and (61) respectively.

(60)   *In Ebron* **regnedØ** *first kyng Dauid .vij. ʒeer & a half*

'In Hebron reigned David, the first king, for four years and a half.' (*Mandeville's Travels*)

(61)   *And ʒif it* **hadde** *Ryueres & welles & the lond also were as it is in oper partyes it scholde ben als full of peple*

'And if it had rivers and wells and the land were as in the other parts, it should have been just as full of people.' (*Mandeville's Travels*)

Additionally, the regression analysis reveals a frequency effect for past singular forms, implying that high-frequency verbs are preferred with zero-endings. This ties in with the idea that high frequency is associated with reductive change (see e.g. Krug 2003;

Bybee 2006). To conclude, the case of the past singular provides more evidence for French loan verbs in Late Middle English being subject to a markedness bias.

## 3.4. Discussion

In this chapter we have demonstrated the existence of loan word accommodation biases in two different contact settings: English loan verbs in Present-day Dutch (cf. De Smet 2014) and French loan verbs in Late Middle English. Analysing data from the CGN, the Helsinki Corpus, and *The Parson's Tale* and *Mandeville's Travels*, we have for both cases identified a syntactic bias (*non-finite bias*) as well as a morphological bias (*markedness bias*). Under the non-finite bias, on the one hand, loan verbs are generally biased towards non-finite forms (e.g. gerund, infinitive, past participle, present participle) as opposed to finite forms (e.g. imperative, past, present). The markedness bias, on the other hand, refers to loan verbs being biased towards less explicitly marked (or simply unmarked) inflectional categories (e.g. single-marked categories) as opposed to explicitly marked inflectional categories (e.g. double-marked categories). Admittedly, there exist slight differences between the contact settings: for example, in Late Middle English the past participle is more frequent than the infinitive, whereas in Dutch the infinitive is more frequent than the past participle. However, those subtle differences do not affect the general findings, which are remarkably similar for both cases.

The empirical evidence for the non-finite bias ranks consistently stronger than for the markedness bias, where evidence is less consistent. By means of illustration, for Late Middle English, French verbs are avoided with zero-inflection in the past singular (e.g. *regnedØ*), but the evidence for a possible preference for zero-endings in present plurals and infinitives is tentative at most. The evidence for loan verbs being biased towards non-finite forms is much stronger. Additionally, the non-finite bias also has intrinsically more pervasive characteristics than the markedness bias. For instance, consider the case of the infinitive and the past participle, which language users exhibit a high preference for under the non-finite bias. However, both forms are inflectionally marked, the clearest example being the past participle in Dutch, which is formed by means of circumfix *ge⟩...⟨t*, and which should thus be avoided following markedness bias rules. Still, the findings for the Dutch and Late Middle English data have shown that such inflected non-finite forms are favoured despite their inflectional endings. In other words, the moment when the biases enter into conflict it is generally the syntactic non-finite bias which dominates over the morphological markedness bias. Similar accommodation biases for adjectives will be presented in Chapter 4., where the syntactic bias also ranks consistently stronger than the morphological bias.

Since accommodation biases are at work in the two contact settings under investigation, we may conclude that even direct insertion — the most commonly used

loan word accommodation strategy — is more differentiated than it seems at first sight. Verbs entering into their recipient language by means of direct insertion can in theory *directly* accommodate to that language and immediately become fully functional. However, this chapter has shown that loan verbs — even under direct insertion — are biased towards specific categories. We may, therefore, conclude that direct insertion is not fully without constraints. In fact, Wohlgemuth's (2009: 291) reasoning that — because of the prevalence of direct insertion worldwide — borrowing is not constrained by inflectional endings should be revised. Admittedly, inflection is not the main obstacle to loan word integration, as we have shown that the constraints are more of a syntactic nature, but it is still an obstacle.

To the best of our knowledge, loan word accommodation biases have not been described this explicitly before, and never under this name, which makes this claim quite innovative. However, several authors have already hinted at (parts of) this claim in previous research. As mentioned in Section 3.1, a similar phenomenon has already been observed for loan verbs from Kriol in Jaminjung, which tend to enter the class of uninflected verbs instead of the inflected verbs (Schultze-Berndt 2017: 265). Also, Harris & Campbell (1995: 135) have already provided some pieces of the accommodation bias puzzle when describing that verbs are less borrowable than nouns because they have more morphosyntactic markings than nouns, which complicates the process of loan word accommodation. This implies that language users are keen to avoid morphosyntactically complex structures when accommodating loan words, which could be linked to avoiding a processing cost in loan word integration, as will be described in depth in Section 5.3.4.

One should be aware when overgeneralising these findings, which apply to two specific contact settings with similar language pairs (Anglo French, Late Middle English, Present-day English, and Present-day Dutch). Not only are they genetically related, they also show many similarities as compared to typologically more distant languages. The recipient languages in both cases, namely Present-day Dutch and Late Middle English, show great typological similarity. Despite their differences relating to the historical context and the intensity of contact, the present case study only represents a minor fraction of all language contact situations worldwide. Whether — and if so, to what extent — the findings can be applied to other contact settings is to date still unclear. In P'urhepecha, for example, Spanish verbs are inserted in their non-finite form, and then an integrating vowel /i/ is added in order to ensure that the form meets the morphophonological requirements of its recipient language (Bellamy 2022 p.c.). Only then inflectional material can be added. However, the present findings align with the findings and theories of several other authors (cf. supra), and we do not see any reasons as to why the underlying linguistic theory should not apply to other (types of) language contact settings as well. In fact,

preliminary findings on Old Norse loan verbs in Late Middle English have shown that Old Norse loan verbs are subject to slight loan word accommodation biases (Elter & Shaw 2022), but this topic will be expanded on in Section 8.2. Also, as we have mentioned above, evidence for a markedness bias in two non-Indo-European languages in contact has already been provided by Schultze-Berndt (2017: 265), who observed that Kriol loan verbs entering into Jaminjung are "integrated in the class of Uninflecting Verbs rather than the class of Inflecting Verbs". Whereas Jaminjung is a Northern Australian language, Kriol is the most frequently used indigenous language in Australia and a *lingua franca*, and the language profiles of both languages are clearly distinct from the ones investigated in this thesis. Another caveat is that authors who exploited non-finite constructions to accommodate French loan words may also have had a tendency to make more use of the same constructions for native English verbs. Therefore, it may be useful to test whether texts with a high density of French loan verbs, such as *Mandeville's Travels* and *The Parson's Tale*, show higher frequencies for any particular non-finite construction type.

An important aim of this thesis is to encompass several potential theoretical explanations for loan word accommodation biases, which will be the emphasis of Chapter 5. However, we already provide an explanation with particular regards to verbs, which tend to have a particularly critical and complex role in clausal syntax. That central role proper to verbs has previously been illustrated by Matras (2009: 182), who asserts that

> verbs accomplish, functionally speaking, two separate things. They are lexical signifiers that label events, activities, states; and they also carry out the grammatical operation of anchoring the predication in the context of the utterance.

Apart from occupying a central role in the clause, verbs typically 'ground', 'instantiate', and 'quantify' events, as described by Langacker (1987). Since verbs have so many different syntactic functions to perform at the same time, they are a particularly complex part of speech for loan words to enter into. While all verbal categories carry this cognitive load, it is even more so for finite verb forms, which tend to carry explicit inflectional marking. As a result, language users may subconsciously try to reduce that processing cost in loan verbs, which is a possible explanation as to why loan verbs tend to be subject to non-finite and markedness biases (cf. Section 5.3.4.).

The above theoretical explanation points to the multiple syntactic functions of verbs as a reason for the existence of accommodation biases in loan verbs. The findings raise the question whether loan word accommodation biases also manifest themselves in other parts of speech than verbs. Also, the effect of French-origin verbs in Middle English being slightly less likely to receive certain inflectional markers was not strong. This is rather surprising, since much of the literature maintains that adding inflections

to loan words is the greatest barrier towards loan word integration (e.g. Harris & Campbell 1995: 135; Sijs 2005: 56–57). Therefore, the next chapter will present a parallel study on loan word accommodation biases in the same two contact settings, but focusing on adjectives.

# Chapter 4

# Accommodation biases in adjectives[43]

## 4.1. Introduction

The previous chapter has provided empirical evidence for the existence of loan word accommodation biases in English loan verbs in Present-day Dutch and French loan verbs in Late Middle English. Although loan verbs and native verbs generally have the same usage potential, loan verbs turned out to occur more in specific usage categories than in others. Since those loan word accommodation biases were linked to the functional complexity of verbs, the question arose whether such biases then appear solely in the category of verbs, or whether they can also appear in other parts of speech. The present chapter will, therefore, present another case study on loan word accommodation biases, this time in the category of adjectives.

Adjectives typically modify nouns and pronouns, but they can also be used adverbially, i.e. when modifying other adjectives. The category is generally characterised as open (Muysken 1981), and — although its borrowability has not been discussed as extensively as that of nouns and verbs (Stammers & Deuchar 2012) — most hierarchies of borrowability present adjectives as being more borrowable than verbs (e.g. Whitney 1881; Muysken 1981; Mugglestone 2006: 74; Winford 2010: 178). At a morphosyntactic level adjectives have complex internal inflectional morphology, since both gender and number can be marked (e.g. Poplack, Sankoff & Miller 1988), but this feature is language-dependent. In Present-day English, for instance, gender and number are not marked. A language where morphosyntactic agreement is particularly transparent is Spanish, since inflectional markings for gender and number are rendered explicitly. In Spanish, grammatical gender is either feminine, as in examples (62) and (64), or masculine, as in (63) and (65).

---

[43] This chapter is based on a manuscript accepted for publication in the journal *Neuphilologische Mitteilungen* (Shaw & De Smet in press). Section 4.4. is based on a case study in De Smet & Shaw (subm.).

(62) *Una <u>mujer</u> **inglesa** logró tener a su primera hija gracias a un tratamiento experimental con esteroides*

'An English woman managed to have her first daughter thanks to an experimental treatment with steroids.' (Corpus del Español: Web/Dialects)

(63) *Un <u>hombre</u> **Inglés** [sic] se quedó atónito al ver que estaban haciendo un periódico diario spiritual*

'An Englishman was astonished when he saw that they were making a spiritual daily newspaper.' (Corpus del Español: Web/Dialects)

(64) *En 1918, se permitió que las <u>mujeres</u> **inglesas** mayores de 30 años pudieran votar*

'In 1918, English women older than 30 years old were granted the right to vote.' (Corpus del Español: Web/Dialects)

(65) *a las tres de la tarde del 27 de junio de 1806, vi entrar 1560 <u>hombres</u> **ingleses***

'At three in the afternoon on 27 June 1806, I saw 1,560 Englishmen enter.' (Corpus del Español: Web/Dialects)

In example (62), adjective *inglés* ('English') is inflected for the feminine singular (*inglesa*) by means of an -*a*-suffix. Example (63), in contrast, illustrates *inglés* being used in a masculine singular form, since it does not receive an additional suffix (∅). The plural of both forms is indicated in examples (64) and (65), where an additional -(*e*)*s*-suffix is added onto the feminine and masculine forms of the singular, hence *inglesas* and *ingleses*.

Adjectives are dependent on nouns, and they can occur in various syntactic positions within a clause. The most typical adjectival positions are attributive (e.g. example (66)), where the adjective precedes its nominal head, and predicative position (e.g. example (67)), where the adjective modifies a nominal head, but is used with a copular verb. A less common syntactic position is postposition (e.g. example (68)), where the adjective follows the nominal head it modifies.

(66) *so I did a thing **last** <u>night</u>, I finally started writing my **first** <u>book</u>!* (Twitter, 13 June 2022)

(67) *We <u>are</u> **happy** to inform that we have all of our calls open now!* (Twitter, 11 June 2022)

(68) *Early admissions to Nottingham General Hospital were reported in the local press, including full lists of the <u>men</u> **concerned**.* (Twitter, 22 November 2020)

Prototypically predicative adjectives tend to align with verbs — more so than prototypically attributive adjectives, which align with nouns (Hollmann 2020: 3).

Indeed, there exist some deep-rooted parallels between predicative adjectives and non-finite verbs. First, as its name suggests, predicative adjectives are *predicates*, which implies that they bear verbal properties. Predicative adjectives are typically combined with a copula, which serves as an auxiliary verb carrying the grammatical information, for instance about the subject (e.g. person). A similar phenomenon can be observed in non-finite verbs, where integration is realised by means of auxiliary or modal verbs. This resemblance between predicative adjectives and non-finite verbs is illustrated in examples (69) and (70) respectively.

(69)   *hij <u>is</u>* **getraind**

      'he is trained'

(70)   *hij <u>heeft</u>* **getraind**

      'he has trained'

In both examples, the grammatical information underlying *hij* (i.e. third person singular) and the temporal domain of the sentence (i.e. present domain) are expressed periphrastically, namely by auxiliary verbs *is* and *heeft*, whereas *getraind* provides the lexical information. Non-finite verbs had already been linked to Wohlgemuth's (2009) light verb strategy in Section 3.2.2., since — in both non-finites and light verb constructions — an auxiliary, modal, or light verb carries the grammatical information and, therefore, the inflections. This feature has now also been found to apply to predicative adjectives. The second similarity between predicative adjectives (71) and non-finite verbs (72) concerns their position in the clause, as both tend to be used clause-finally.

(71)   *De appel is* **heerlijk**.

      'The apple is delicious.'

(72)   *Hij heeft een appel* **gegeten**.

      'He has eaten an apple.'

In contrast, attributive adjectives (73) and finite verbs (74) tend to occur in clause-medial position.

(73)   *Het is een* **heerlijke** *appel*.

      'It is a delicious apple.'

(74)   *Hij* **eet** *een appel*.

      'He eats an apple.'

This clause-medial position implies that language users — when they use a loan word or code-switch as an attributive adjective or a finite verb — need to switch languages within a single clause. This is different for predicative adjectives and non-finite verbs, where clause-final position generally does not require language users to switch languages within the clause. The code-switching literature has indeed hinted at switching being less costly in clause-final than in clause-medial position (e.g. Flamenbaum 2014: 351; Amuzu 2016: 138). This will be expanded on in Section 5.5.

Based on the above similarities between predicative adjectives and non-finite verbs, we hypothesise that loan adjectives — like loan verbs — will be subject to probabilistic accommodation biases, and more concretely at two levels. First, we hypothesise that loan adjectives will be biased towards predicative syntactic position as opposed to attributive position. This syntactic tendency will be referred to as the *predicative bias*. Although attributive syntactic position is generally more frequent (e.g. in Middle English, see Burrow & Turville-Petre 1992: 44–45), predicative syntactic position is generally less complex[44]. This became clear from the two resemblances with non-finite verbs, and from first-language acquisition research, which has shown that children acquire predicative use before they acquire attributive use (cf. Section 5.3.1.). Second, we hypothesise that loan adjectives will be biased towards inflectionally unmarked forms as opposed to inflectionally marked forms[45]. Similar to verbs, this morphological bias will be referred to as the *markedness bias*. For both parts of speech, the markedness bias refers to a preference of language users to use loans without any type of inflection. This hypothesis does not only build on the evidence of the markedness bias as found for loan verbs, but also on the corresponding evidence in the literature for loans being easier to integrate in categories with fewer inflectional markings (e.g. Harris & Campbell 1995; Schultze-Berndt 2017: 265). Furthermore, the markedness bias is a logical extension of the predicative bias, since predicative adjectives in the contact settings under investigation are inflected less often than attributive adjectives.

We will now turn to the evidence, which is presented in two complementary corpus studies on the same two contact settings as for verbs, i.e. the English-Dutch and French-Middle English contact settings. The settings have intentionally been kept constant. In both contact settings, loan adjectives can generally be integrated by means of direct insertion (cf. Wohlgemuth 2009), as is illustrated in (75) for the English-Dutch contact setting, and in (76) for the French-Late Middle English

---

[44] This does not mean that non-finite verb forms and predicative adjectives cannot be morphosyntactically complex. An example is the *for to*-infinitive in English, as in *for to learn*, and the past participle, which is formed by means of a circumfix.

[45] Forms being 'marked' refers to marking being added to the stem.

contact setting (also see Sections 2.4.3.3. and 2.4.4.3.). In both cases, recipient-language *-e* is attached to the loan adjective (*fair-e* and *round-e* respectively).

(75)  *wij willen altijd uh onderhandelen maar dat moet op een **faire** basis gebeuren*

'We are always willing to, um, negotiate but it has to be done on a fair base.' (CGN)

(76)  *The clerkes han **rounde** crounes*

'The clerks have round crowns.' (*Mandeville's Travels*, c1400, PPCME2)

Note that we are well aware that the term 'direct insertion' traditionally refers to contexts where loan *verbs* are accommodated into a recipient language, and that it does not refer to loan adjectives in Wohlgemuth's (2009) definition (cf. Section 2.4.3.3.). However, Wohlgemuth (2009: 90, 95) also uses the term to refer to action-word nouns which resemble verbs at a semantic level, and to nouns which are used verbally by means of a verbalizer. This indicates that Wohlgemuth's (2009) definition is not so strictly limited to verbs, and that we can use the term 'direct insertion' for adjectives as well. The mechanism for adjectives is identical to the mechanism for verbs: recipient-language inflections can be added directly onto loan (adjective) stems.

Whereas Section 4.2. presents the data for English-origin adjectives in spoken Present-day Dutch (extracted from the CGN), Section 4.3. presents the data for French-origin adjectives in Late Middle English (extracted from the PPCME2). For both case studies this chapter first provides a detailed description of the used data and the corresponding methodology, followed by an overview of the findings, which provide empirical evidence for the existence of a predicative and markedness bias. The data for Middle English additionally reveal a head bias, meaning that French-origin adjectives — as compared to English-origin adjectives — are more likely to occur with French-origin nominal heads. The existence of accommodation biases in loan adjectives raises yet another question, namely how such biases tend to evolve through time. This expansion is the topic of Section 4.4., where biases traced in French loan adjectives in Early and Late Middle English are shown to be remarkably persistent, and — in the case of the markedness bias — even stronger for earlier than for later loans. Section 4.5. extensively discusses the findings for this chapter and ends with some conclusions.

## 4.2. English-Dutch contact setting

### 4.2.1. Data and methodology

#### 4.2.1.1. Data extraction

In order to conduct this case study on loan adjectives, we followed the methodology used for loan verbs in Chapter 3. wherever possible. As such, we started from a list of English items marked as adjectives in the BNC (Bodleian Libraries 2007), which were then searched for in the CGN (Nederlandse Taalunie 2004). This set of English adjectives served as the loan set and was paired to a control set containing native Dutch adjectives, but which may also contain a few older, well-established loan adjectives from, for instance, Latin and Greek. This control set was a random sample of 2% of all Dutch adjectives in CGN. Whereas the loan set contained 447 relevant attestations, the control set contained 830 relevant attestations. This means that — as opposed to the case study on loan verbs — the loan and control set were not frequency-matched. A total of 1,277 relevant attestations were retained in the dataset, which includes both high- and low-frequency words.

A number of attestations were excluded from the dataset. First of all, we excluded adjectives used with an intensifying prefix (e.g. *kei-* 'boulder', *reuze-* 'giant', *super-* 'super'), fixed expressions (e.g. *zeker en vast* lit. 'sure and fixed', meaning 'definitely'), lexicalised elements (e.g. *bitter lemon*), abbreviations (e.g. *aso*, from *asociaal* 'antisocial'), and the lemma *half* 'half' as in *half tien* (lit. 'half ten', meaning 'half past nine'). Second, we did not include adjectives which fall outside of the envelope of variation between attributive and predicative syntactic position: whereas some adjectives can exclusively be used predicatively (e.g. *allang* 'already, by now', *overstuur* 'upset'), others can exclusively be used attributively (e.g. *gans* 'entire, whole', *luxe* 'luxury'). Adjectives which are mainly used adverbially (e.g. *eventueel* 'possibly', *hopelijk* 'hopefully', *waarschijnlijk* 'probably') were not included either. Third, we excluded comparatives and superlatives, and (adverbial) adjectives which are part of separable infinitives, such as *vasthouden* 'hold on'. Fourth, the loan set does not include any cognates with Dutch (e.g. *wild* 'wild'), nor English homographs of existing Dutch adjectives, such as *glad*, which can be a native Dutch adjective meaning 'slippery' as well as an English loan adjective meaning 'happy'. Loans with more than one possible source language were excluded for the sake of transparency in the data analysis process. Last, we excluded adjectives phonetically ending in a vowel[46] (e.g. *extra*, *happy*, *heavy*) on account of Saugera's (2012) findings for English loan adjectives in Present-day French: English adjectives ending in *-y* are generally not accommodated to the French inflectional system. Although she does not mention the vowels other

---

[46] English loan adjectives ending in vowels are categorically not inflected in Dutch either.

than *-y* being obstructive to regular loan word accommodation, we decided to be careful and to exclude those as well.

### 4.2.1.2. Data annotation

The CGN provides automatic annotations, which were manually verified, expanded and — in some cases — simplified, and we added manual annotations for adjectival syntactic position and inflectional form. For syntactic position, first, we distinguished between attributive, non-attributive and other positions. The category of 'other positions' in our data refers to the following three subtypes: (i) adverbial adjectives, which modify a (non-copular) verb, adverb, or another adjective, (ii) nominal adjectives, which are used nominally, and (iii) sentence adverbials, where the adverbial puts a scope over the entire sentence. Examples (77) and (78) illustrate adverbial adjectives, the adjective in (77) being used with a verb (*gekend*), and the one in (78) with an adverb (*goed*). Example (79), next, presents two nominally used adjectives, namely *goeie* and *enige*, which are both used with article *de. Moeilijk* in (80) is an example of a sentence adverbial. All examples are in Dutch.

(77)     *ik heb mijn voorganger zelf uh niet* **persoonlijk** *gekend* .

        'I didn't know, um, my predecessor personally.' (CGN)

(78)     *maar dit kon ik* **aardig** *goed uh ja kon 'k aardig goed volgen* .

        'But this I followed pretty well, um, yes, I could follow it pretty well.' (CGN)

(79)     *en ik heb mijn methodes gebruikt . – en ik denk dat dat de* **goeie** *waren dat dat de* **enige** *waren* .

        'And I have used my methods . – and I think that they were the right ones that they were the only ones.' (CGN)

(80)     *op dit ogenblik rijden ze nog allemaal uh de favorieten toch vooraan gegroepeerd een beetje nog ge- geen grote afscheiding maar dat kan ook* **moeilijk** *want we zijn nog maar uh tien minuten weg*

        'At the moment they are still all riding, um, the favourites grouped together in the front a bit n- no big separation yet but that would be difficult (lit. that can difficult) because we left only ten minutes ago.' (CGN)

However, under the scope of this thesis only attributive and non-attributive positions were included in the data analysis. An example of attributive adjectival position in Dutch, where the adjective (*klein*) precedes the head (*voortuintje*) it modifies, is provided in (81).

(81)   *hij heeft alleen maar zo'n* **klein** <u>*voortuintje*</u>

      'He only has a kind of small front yard.' (CGN)

Non-attributive adjectival position in our data annotations covers three subtypes: (i) predicative position, where the adjective is used with a copular verb, (ii) postposition, where the adjective follows the head it modifies, and (iii) secondary predicates, where the adjective is predicated of the patient argument of a higher verb. Examples for those three different subtypes are presented in (82), (83), and (84).

(82)   *je hebt gezegd het* <u>*is*</u> **klein** *en* **handig** *maar je bent nog één voordeel vergeten* .

      'You have said it is small and useful but you have forgotten about one more advantage.' (CGN)

(83)   *je ziet toch wel* <u>*iets*</u> **bijzonders***?*

      'You see something special, don't you?' (CGN)

(84)   *nee je kiest* <u>*Frans*</u> *omdat je't* **leuk** *vindt* .

      'No you choose French because you like it (lit. you find it fun).' (CGN)

In case of ellipses, which can complicate the identification of syntactic position, the context was taken into account, but when a sufficiently large amount of contextual features was lacking those cases had to be excluded from the dataset.

Apart from syntactic position, the attestations were also annotated for their inflectional forms. For the attestations where inflectional endings are to be expected, we verified whether those endings were congruous to the endings actually attested. Notice that not all adjectives are part of the envelope of variation between inflection and non-inflection: whereas some adjectives only exist with invariable *-e* (e.g. English loan adjective *safe* and native Dutch adjectives *geachte* ('dear' in formal usage) and *zelfde* ('same')), other adjectives can exclusively be used without *-e* (e.g. *eender* 'any'). The Dutch language also contains a number of adjectives for materials, which invariably end in *-en*, as in *glazen* 'glass', *houten* 'wooden', and *ijzeren* 'iron'. Such lemmas could not be included in the data analysis either. Table 13 is based on the *Algemene Nederlandse Spraakkunst* (E-ANS; Coppen, Haeseryn & de Vriend 2002) and presents an overview of the adjectival inflectional system as used in Present-day Dutch, exemplifying *klein* ('small')[47]. It shows how the article and inflectional endings generally differ based on syntactic position, number, and gender of the head (common[48] and neuter). Dutch articles can be definite (*de, het*) or indefinite (*een*), and

---

[47] We are aware that loan adjectives may not always be inflected exactly like native adjectives, but this table can serve as a base.

[48] 'Common gender' entails both feminine and masculine.

inflectional endings can be compulsory, impossible, or optional. Whereas predicative adjectives in Dutch are never inflected, attributive adjectives are inflected, with the exception of adjectives modifying neuter singular nouns in an indefinite noun phrase. This is different from English, where adjectives — regardless of their syntactic position — cannot be inflected.

*Table 13: Overview of the adjectival inflectional system in Present-day Dutch.*

| SYNTACTIC POSITION | NUMBER | GENDER | DEFINITENESS | INFLECTION | EXAMPLE |
|---|---|---|---|---|---|
| Attributive | Plural | | | Compulsory | *klein-e* |
| | Singular | Common | | | |
| | | Neuter | Definite | Optional | *klein(-e)* |
| | | | Indefinite | Impossible | *klein* |
| Non-attributive | | | | | |

Although this overview accounts for the large majority of Dutch adjectives, it is not entirely fool-proof because of two reasons. First, the Dutch language omits inflectional endings in cases where the adjective refers to the exceptionality of a person (Onze Taal 2022). This is, for example, the case in *een **groot** man* 'a great man', which is a common form, and which should, therefore, in theory be inflected. However, in this case, the absence of inflection stresses to what extent the man is said to be morally or intellectually exceptional. Inflected *een **grote** man*, in contrast, has a different and more literal meaning, namely 'a tall man'. However, the group of adjectives in which inflectional endings are omitted in order to realise a semantic difference is negligible and should not pose a problem for our data analysis. Second, adjectival inflections may slightly diverge in some dialectal attestations, as in examples (85) and (86).

(85)   **klein** *beddekes*

'small beds' (CGN)

(86)   *in zijne **vrijen** tijd*

'in his spare time' (CGN)

Since *klein* in (85) is used in an indefinite attributive in the plural, it should be inflected in Standard Dutch: *kleine*. *Vrij* in example (86) should in Standard Dutch be inflected as *vrije*, but receives an additional *-n*-suffix, hence *vrijen*. However, this should not raise

issues since the total number of dialectal attestations in the dataset is limited to four attestations, and they are indicated as such, so that any potential effects originating from dialectal inflections can easily be detected.

### 4.2.1.3. Data analysis

Visualisations for this corpus study were realised in R, using the packages "ggmosaic" (Jeppson et al. 2021) and "ggplot2" (Wickham 2016). The use of ggplot2 has been described in Levshina's (2015) ninth chapter on the associations between categorical variables. By means of the Fisher's exact test (for a description, see Section 3.2.1.3.), we additionally calculated p-values for the different categories: attributives or non-attributives, and — if applicable — with or without inflection. Significance was set at $p < 0.05$ as significance threshold.

## 4.2.2. Findings

The data were first visualised in a mosaic plot, which represents the behavioural distribution of English-origin adjectives and control adjectives in the following four categories:

1. Non-attributive adjectives (abbreviated as "Non-attr.")
2. Attributive adjectives in contexts where inflection is optional (abbreviated as "Attr.infl.option.")
3. Attributive adjectives in contexts where inflection is compulsory (abbreviated as "Attr.infl.compuls.")
4. Attributive adjectives in contexts where inflection is impossible (abbreviated as "Attr.infl.imposs.")

Syntactic position and inflectional endings were combined into those four categories. The vertical divisions in Figure 7 represent the frequency distributions of English-origin and control adjectives in the four above categories, whereas the width of the horizontal bars reflects the number of observations in each usage category. The vertical dashed line in Figure 7 is set at 35%, separating English-origin adjectives (447) from Dutch-origin control adjectives (830), and showing what a hypothetical 'bias-free' distribution in the dataset would look like.
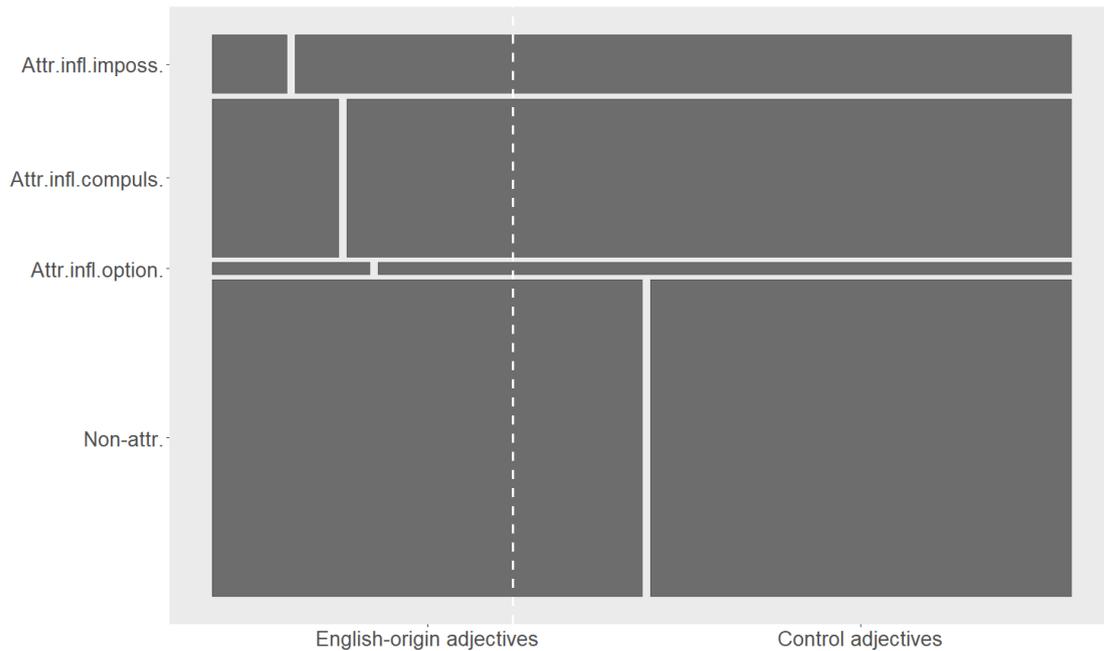
*Figure 7: Distribution of syntactic position and inflection for English-origin adjectives and control adjectives (n = 1,277) with vertical dashed line set at 35%.*

The vertical divisions in Figure 7 suggest that the distribution of loan and control adjectives in the different categories is subject to biases. A clear behavioural difference can be observed for syntactic position: starting from the bottom, the plot shows that English-origin adjectives are overrepresented in non-attributive position, whereas control adjectives are overrepresented in attributive position. The relative frequencies reported in Table 14 corroborate this observation, since 83.9% of the attestations for English loan adjectives occur in non-attributive position, whereas this is only the case for 44.2% of the adjectives in the control set. With $p < 0.001$ this effect is strongly significant.

*Table 14: Distribution of English-origin adjectives and Dutch-origin adjectives in attributive and non-attributive syntactic position (Fisher, $p < 0.001$).*

|  | Attributive position | Non-attributive position |
|---|---|---|
| Dutch-origin adjectives (n = 830) | 463 (55.8%) | 367 (44.2%) |
| English-origin adjectives (n = 447) | 72 (16.1%) | 375 (83.9%) |

As hypothesised, English loan adjectives in Dutch are thus biased to non-attributive syntactic position, as in example (87). This finding is in stark contrast with control adjectives, which are disproportionally more frequent in attributive syntactic position, as in example (88).

(87)   *is <u>is</u> ze zo* **close** *met Chris dan dat zij ook op de verjaardag komt blijkbaar ?*

     'Is-is she that close to Chris then that she also visits for the birthday apparently?' (CGN)

(88)   *van Bratislava tot Barcelona overal wordt gewerkt aan ecotopia's aan ecopolis en* **verwante** <u>*begrippen*</u> .

     'From Bratislava to Barcelona everywhere they are constructing 'ecotopias' and 'ecopoles' and related terms.' (CGN)

The observations for inflectional form in Figure 7 are less pronounced than the ones for syntactic position, since the differences among the three categories of attributive adjectives (*-e*-inflection, zero-inflection, or optional *-e*-inflection) are highly subtle. Table 15 provides concrete numbers for English-origin adjectives and control adjectives in the three inflectional categories.

*Table 15: Distribution of attributive English-origin adjectives vs. Dutch-origin adjectives with -e-inflection, zero-inflection, or optional -e-inflection (Fisher, p = 0.88).*

|  | *-e* | Zero | Optional *-e* | |
|---|---|---|---|---|
| Dutch-origin adjectives (n = 463) | 316 (68.3%) | 125 (27%) | 22 (4.8%) Inflected 19 (86.4%) | Uninflected 3 (13.6%) |
| English-origin adjectives (n = 52) | 35 (67.3%) | 12 (23.1%) | 5 (9.6%) Inflected 0 (0%) | Uninflected 5 (100%) |

As becomes clear from Table 15, 68.3% of the Dutch-origin adjectives appear in contexts with *-e*-inflection, as compared to only 67.3% of the English-origin adjectives. However, the difference between adjectives of both source languages is negligible. In contexts with zero-inflection, Dutch-origin adjectives (27%) are again more frequent than English-origin adjectives (23.1%). However, in cases where inflection is optional English-origin adjectives (9.6%) are more common than

Dutch-origin adjectives (4.8%). This means that the findings run in the expected direction, namely that English loan adjectives tend to be avoided in contexts where inflection is required. In example (89), for instance, loan adjective *relaxed* is used in an indefinite neuter singular form, where it is not inflected.

(89)    *uh ben hun dus tegengekomen uh en uhm ja op zich best best een* **relaxed** *uh stel eigenlijk* .

    'Um so I ran into them, um, and, um, yes, actually quite-quite a relaxed, um, couple actually.' (CGN)

However, the findings for this markedness bias are not proven significant according to the Fisher's exact test (p = 0.88), where contexts with *-e*-inflection are compared to contexts with zero-inflection and optional *-e*-inflection. The reason why categories with zero-inflection and optional *-e*-inflection were taken together in the significance test is that — in both contexts — schwa is or can be deleted. The lack of significance may be due to the differences in the proportions not being sufficiently large, or to the sample size being too small.

Finally, zooming in solely on those contexts where inflection is optional (i.e. right column in Table 15), it is remarkable that 86.4% of the Dutch-origin adjectives in this context are inflected, whereas none of the English-origin adjectives are inflected (0%). The effect is strongly significant with p < 0.001, but it should not be generalised due to the small number of attestations. Still, the fact that English loan adjectives in Present-day Dutch are — at least to some extent — avoided in contexts where inflection is optional indicates again that a markedness bias in this contact setting should not be rejected.

In Section 4.3. we will ascertain whether the predicative and markedness biases are attested in the French-Late Middle English contact setting as well, or whether they are solely a feature of English loan adjectives in Dutch.

## 4.3. French-Late Middle English contact setting

### 4.3.1. Data and methodology

#### 4.3.1.1. Data extraction

This corpus study presents a close analysis of three prose texts, which have all been written in the late 14th century: *The Parson's Tale* (c1390), the Wycliffite version of *The Old Testament* (c1398), and *Mandeville's Travels* (c1400). Whereas *The Parson's Tale* and *Mandeville's Travels* have already been presented in Chapter 3., *The Old Testament* has not been used in the previous analyses. The main reason for adding this text here was to create a larger body of texts, which was less essential in the case study on verbs, where the dataset included *The Parson's Tale* and *Mandeville's Travels* as well as the M3 period from the HC. *The Old Testament* is a biblical text, included in the Wycliffite Bible

and translated from the Latin vulgate. Like *The Parson's Tale* and *Mandeville's Travels* it has been written up at the time of the borrowing peak, when the number of French loans in Late Middle English surged drastically. The three texts have been written up in the East Midlands dialect, which was "the largest and most populous of the major dialect areas" (Baugh & Cable 2013: 187). The text editions used in this case study were those included in the second version of the PPCME2 (Kroch & Taylor 2000), which has been presented in detail in Section 3.3.1.1.1. Note that we analysed the full texts of *The Parson's Tale* and *Mandeville's Travels*, but only a 9,910-words sample for *The Old Testament*. This text sample was extracted from the Helsinki Corpus (Rissanen et al. 1991).

We excluded some of the attestations parsed as adjectives in the MED (Lewis 1952–2001). Those include grammaticalised items such as *all* (90), *such* (91), and *thilk* (92), which tend to occur as determiners (e.g. *any*, *some*) or predeterminers (e.g. *all*, *both*) and which, therefore, do not show adjective-like behaviour.

(90)   *For it is the herte and the myddes of **all** the world*

       'Because it is the heart and the middle of the whole world.' (*Mandeville's Travels*)

(91)   *it is wel a .iij. iourneyes of **suche** weye to passe from Prusse to the lond of sarazin habitable .*

       'It is a four-day trip of such length to go from Prussia to the habitable land of the Saracen.' (*Mandeville's Travels*)

(92)   *ne nothyng is to hym moore abhomnyable than **thilke** milk whan it is medled with oother mete .*

       'Nothing is more abominable to him than such milk when it is mixed with other food.' (*Mandeville's Travels*)

After having excluded the above attestations, we took a 50% subset of our sample, and retained 1,887 attestations[49], of which 1,601 are of English origin (84.9%) and 286 are of French origin (15.1%). The total number of adjectival tokens was 294. The subset included both high- and low-frequency items, since both types can provide critical insights into the process of loan adjective accommodation.

### 4.3.1.2. Data annotation

Adjectival attestations in the subset were manually lemmatised and annotated for source language, adjectival syntactic position, and inflectional form. For source language, a main distinction was drawn between French-origin and English-origin adjectives, and this classification was based on the etymological information as provided in the MED (Lewis 1952–2001). Some examples of French-origin adjectives

---

[49] This is the number including nominal and adverbial adjectives.

are *innocent*, *horrible*, and *pacient* 'patient', whereas examples of English-origin adjectives are *dedli* 'deadly', *god* 'good', and *wild*. It should be noted that the set of French-origin adjectives includes any attestations which can possibly be linked to French origin, including adjectives of joint French and Latin origin. However, adjectives of exclusively Latin origin (e.g. *corrupt*, *delicate*, *disordinate*) were excluded, and so were adjectives which exceed the divide between English and French origin and are attested as mixed (e.g. *fresh*, *long*, *precious*). Since most adjectives depend on nouns, those head nouns were annotated for source language as well, applying the same annotation scheme as for adjectives. The categorisation of adjectival syntactic position (or function), next, was either attributive (93), predicative (94), adverbial (95), or nominal (96). Following Brunner (1963: 53), the nominal use of adjectives in Middle English was much more common than in later stages of English. Note that adverbial position only includes adjectives which can possibly be used adjectivally, and not, for instance, adverbs such as *enough*, *forth*, and *thider* 'there, to that place'.

(93)  *And men seyn þere þat is a tokene þat the Emperour hath ylost a* **gret** *partie of his londes & of his lordschipes .*

      'And men say there that it is a sign that the emperor has lots a great part of his land and of his lordships.' (*Mandeville's Travels*)

(94)  *Thou shalt considere what thow art that doost the synne , wheither thou* **be male** *or* **femele** *,* **Yong** *or* **oold** *,* **gentil** *or* **thral**

      'You shall consider what part of you is doing the sin, whether you are male or female, young or old, a noble or a peasant.' (*The Parson's Tale*, c1390, PPCME2)

(95)  *And þere he appered* **first** *to Mary Magdalene whan he was risen*

      'And there he appeared first to Mary Magdalene when he was resurrected.' (*Mandeville's Travels*)

(96)  *Also in þat contree & in* **opere** *also men fynden longe Apples to selle*

      'Also, in that country and in others as well, men find long apples to sell.' (*Mandeville's Travels*)

However, only attributive and predicative position were included in the analysis. (Pre)determiners were not included in the annotations. Attributive and predicative adjectives were annotated for their adjectival inflectional form, which took into account the inflectional variation as attested in the sample only. This is because the adjectival inflectional paradigm — like other types of inflection — in Late Middle English was subject to deflection (e.g. Pyles & Algeo 1982: 153–154;

Rissanen 2000)[50], which is exactly when *Mandeville's Travels*, *The Old Testament*, and *The Parson's Tale* have been written. In fact, from all parts of speech the change in the inflectional system of adjectives was the strongest (Brinton & Arnovick 2011: 283). More concretely, this means that the declension system increasingly allowed for inflectionally reduced forms, i.e. forms ending in zero-inflection (cf. the b-options in examples (97) and (98)). Another reason to focus solely on the inflectional endings attested in our dataset is that inflection in Middle English varied among dialects and composition dates, which should not cause issues in our sample since all three texts have been written in the same dialect (i.e. East Midlands dialect) and around the same period (i.e. late 14th century). The adjectives in our sample allow for variation between zero- and schwa-endings in both attributive and predicative position, although inflection is generally less common in predicative position (Brunner 1963: 51). The Old English distinction between strong and weak adjectival declensions had been lost by Middle English. Inflections of grammatical gender as well had been lost by the beginning of the Middle English period (Mustanoja 1960: 43–45). Examples of variation between schwa- and zero-inflection for attributive and predicative adjectives are presented in (97) and (98) respectively.

(97)    a. *For I haue often tymes passed & ryden þat way with* **gode** *companye of many lordes*

'For I have many times passed and ridden that way with [the] good company of many lords.' (*Mandeville's Travels*)

b. *For certes , in the werkynge of the deedly synne , ther is no trust to no* **good∅** *werk that we han doon biforn*

'For sure, in the act of the deadly sin, there is no trust to no good work that we have done before.' (*The Parson's Tale*)

(98)    a. *þei ben* **gode** *& of gret virtue*

'They are good and of great virtue.' (*Mandeville's Travels*)

b. *And God sei3 that it was* **good∅**

'And God saw that it was good.' (*The Old Testament*, c1395, PPCME2)

Apart from zero- and schwa-endings, we also encountered three nominal attestations of *some*, ending in *-en*, hence *summen*. An in-context example is provided in (99).

---

[50] However, there existed a clear revival of the distinction between strong and weak inflections "in the metre of careful writers of the South, such as Gower and Chaucer" (Burrow & Turville-Petre 1992: 29).

(99)   *And* **summen** *seyn þat þei ben sepultures of grete lordes þat weren somtyme*

'And some say that they are tombs of great lords who lived in the past.' (*Mandeville's Travels*)

The sample also contained some adjectival attestations ending in *-(e)s*, which occurs in twelve nominal adjectives used in the plural. We also encountered three attestations of attributive adjectives retaining their French-origin *-e(s)*-endings, which will be illustrated and discussed in depth in Section 4.3.2.2. However, since the number of inflectional forms diverging from zero and schwa is so limited, they were not included in the quantitative analysis. Another type of adjectives which was not incorporated in the analysis are those in invariable *-e* (e.g. *large *larg, reasonable *reasonabl*), since such adjectives do not cover the envelope of variation in inflectional endings. This category only includes adjectives which end in schwa in each possible spelling variant, but not adjectives where inflection can be avoided in one of the spelling variants (e.g. *trouble*, which can be written as *troubel*). One should notice that lemma frequency was not incorporated in the manual annotations, as it was included directly in data analysis.

### 4.3.1.3. Data analysis

Data analysis for this case study consisted in comparing the distributional properties of French-origin and control adjectives in attributive and predicative position, and with and without inflection. This was done by means of a mosaic plot, created with the R-packages "ggmosaic" (Jeppson et al. 2021) and "ggplot2" (Wickham 2016). We added p-values, calculated by means of the Fisher's exact test, for syntactic position and inflectional form. We also set up an additional mixed-effects logistic regression model, using R-package "lme4" (Bates et al. 2015), in order to gain more insight into the findings for adjectival inflectional form. The function used for the model was glmer() rather than lmer(), which is desirable when the outcome of the dependent variable in the model is binomial rather than Gaussian.

## 4.3.2. Findings

First data inspection was done by visualising the data for Late Middle English in the mosaic plot in Figure 8, which illustrates how often French- and English-origin adjectives occur in the six following categories, covering both adjectival syntactic position and inflectional form:

1. Predicative adjectives in contexts where inflection is invariable (abbreviated as "Pred.infl.invar.")
2. Uninflected predicative adjectives (abbreviated as "Pred.uninfl.")
3. Inflected predicative adjectives (abbreviated as "Pred.infl.")
4. Uninflected attributive adjectives (abbreviated as "Attr.uninfl.")

5.  Attributive adjectives in contexts where inflection is invariable (abbreviated as "Attr.infl.invar.")
6.  Inflected attributive adjectives (abbreviated as "Attr.infl.")

Similar to Figure 7, the vertical bars in the mosaic plot reflect the frequency distributions of loan and control adjectives in the above six categories, and the width of the horizontal bars reflects how many observations are attested in each usage category separately. The vertical dashed line shows the distribution of English- and French-origin adjectives and is set at 15.1%. If any biases are attested, the findings will not perfectly coincide with that line.
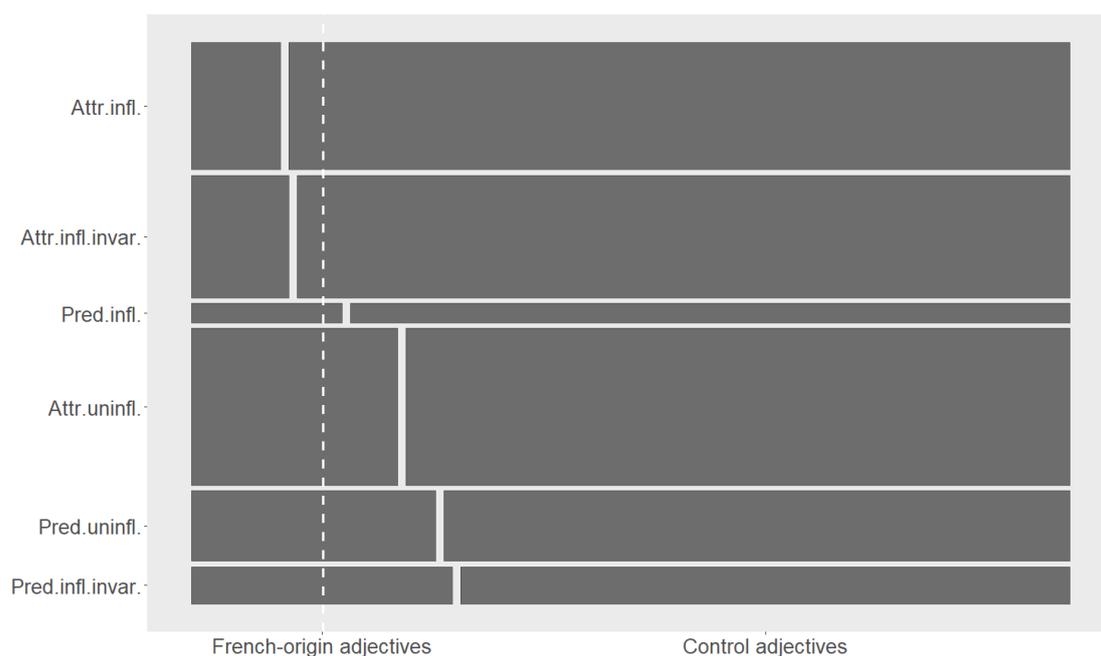


*Figure 8: Distribution of syntactic position and inflection for French-origin adjectives and English-origin adjectives (n = 1,887) with vertical dashed line set at 15.1%.*

At first sight it appears from the vertical divisions that French-origin adjectives are overrepresented in predicative as compared to attributive syntactic position. This confirms the hypothesised predicative bias. Within both predicative and attributive adjectives, French loan adjectives are overrepresented in structures where inflection is invariable and in uninflected structures as compared to inflected structures. Moreover, the only case where a French loan adjective is more frequent in attributive adjective in predicative position is when the attributive adjective is not inflected, and the predicative adjective is. This is an indication of a markedness bias. Both tendencies will be further investigated in what follows.

### 4.3.2.1. Predicative bias

Figure 8 suggested that that French loan adjectives are considerably more common in predicative than in attributive syntactic position. This finding is corroborated by the numbers in Table 16 below, which show that 34.9% of the French loan adjectives are used in predicative position, as opposed to only 21.3% of the native English adjectives. Conversely, native English adjectives are more common in attributive position (78.7%) than French loan adjectives (65.2%). With $p < 0.001$ this finding for syntactic position is strongly significant.

*Table 16: Distribution of French-origin adjectives and English-origin adjectives in attributive and predicative syntactic position (Fisher, $p < 0.001$).*

|  | Attributive position | Predicative position |
|---|---|---|
| English-origin adjectives (n = 1,171) | 922 (78.7%) | 249 (21.3%) |
| French-origin adjectives (n = 264) | 172 (65.2%) | 92 (34.9%) |

An example of this predicative bias is given in (100), where French-origin *pur* ('pure') and *clear* are used predicatively, while English-origin *gret* ('great') in (101) is used attributively.

(100)  *And for als moche as it ne reyneth not in þat contree but the eyr <u>is</u> alwey* **pure & cleer**

'And as long as it does not rain in that country the air is always pure and clear.' (*Mandeville's Travels*)

(101)  *And aboue the cytee is a* **grete** <u>*mountayne*</u> *þat also is clept Polombe*

'And above the city is a great mountain which is also called Polombe.' (*Mandeville's Travels*)

As mentioned in Section 4.3.1.2. predicative adjectives in Late Middle English are typically not inflected (Brunner 1963), which evokes the question whether the bias of French loan adjectives towards predicative position relates to the markedness bias. More concretely, French loan adjectives may be drawn towards predicative structures because inflection is often omitted. This will be addressed in the next section.

### 4.3.2.2. Markedness bias

As to inflection, the mosaic plot in Figure 8 suggested that French loan adjectives are more frequent in uninflected structures and in structures where inflection is invariable than in inflected structures. Although predicative position — regardless of inflectional endings — is overrepresented in loan adjectives, the context in which French-origin adjectives are least common is when predicative adjectives are inflected.

Those findings are now further explored in Table 17 (for attributive position) and Table 18 (for predicative position). Different from Dutch, predicative adjectives in Late Middle English can be inflected, which explains why attributive and predicative adjectives are presented separately. This is the same in Anglo French, where both predicative and attributive adjectives can be inflected, among others for their number (Goyens 2020 p.c.). Table 17 and Table 18 below demonstrate the distribution of the three types of adjectival inflection found in Late Middle English: inflected adjectives, uninflected adjectives, and adjectives in invariable -*e* (e.g. *som*ẹ).

*Table 17: Distribution of attributive French-origin adjectives vs. English-origin adjectives with inflection, without inflection, and with invariable -e (Fisher, p < 0.001[51]).*

|  | Inflected | Uninflected | Invariable -*e* |
|---|---|---|---|
| English-origin adjectives (n = 922) | 307 (33.3%) | 322 (34.9%) | 293 (31.8%) |
| French-origin adjectives (n = 172) | 35 (20.4%) | 100 (58.1%) | 37 (21.5%) |

For attributive adjectives the rate of French-origin adjectives (20.4%) with inflection is remarkably lower than for English-origin adjectives (33.3%); conversely, French-origin adjectives (58.1%) are attested more often in uninflected forms than English-origin adjectives (34.9%). Given p < 0.001 this tendency is strongly significant. Another observation is that French-origin adjectives (21.5%) are less common with invariable -*e* than English-origin adjectives (31.8%). Similar tendencies are detected for predicative adjectives:

---

[51] The p-values in Table 17 and Table 18 depict the distribution of inflected and uninflected forms. Forms in invariable -*e* are not taken into account in the Fisher's exact test.

*Table 18: Distribution of predicative French-origin adjectives vs. English-origin adjectives with inflection, without inflection, and with invariable* -e *(Fisher, p = 0.15).*

|  | Inflected | Uninflected | Invariable -*e* |
|---|---|---|---|
| English-origin adjectives (n = 249) | 43 (17.3%) | 136 (54.6%) | 70 (28.1%) |
| French-origin adjectives (n = 92) | 9 (9.8%) | 53 (57.6%) | 30 (32.6%) |

In predicative position, French-origin adjectives (9.8%) are less common with inflection than English-origin adjectives (17.3%). French-origin adjectives (57.6%) are more common without inflection than English-origin adjectives (54.6%). Although this tendency towards a markedness bias points in the hypothesised direction, the findings are not significant (p = 0.15). This may be due to the dataset being too small for the Fisher's exact test to identify statistical effects. Different from attributive adjectives, French loan adjectives (32.6%) used in predicative position are slightly more common with invariable -*e* than native English adjectives (28.1%).

Two main observations stand out from the numbers given in Table 17 and Table 18. First, French-origin adjectives in Late Middle English are generally avoided with inflection and are thus — as hypothesised — subject to a markedness bias. However, this bias is only significant for adjectives in attributive position. A similar bias has been found for French loan verbs in Late Middle English (Section 3.3.2.2.). Second, the numbers show that predicative adjectives less often carry inflections than attributive adjectives, both in native and loan adjectives, which corroborates the literature on adjectival inflections in Middle English (see e.g. Brunner 1963: 51).

Despite the markedness bias, a substantial number of French loan adjectives can be inflected with seeming ease, as in example (102), where French-origin *vertuous* carries a schwa-ending in attributive position.

(102)   *And þat lond he chees before all oþ londes as the beste & most worthi lond & the most* **vertuous<u>e</u>** *lond of all the world* .

'And he chose that country over all other countries as the best and the most worthy country and the most virtuous country of all the world.' (*Mandeville's Travels*)

As we briefly mentioned in Section 4.3.1.2., French loan adjectives occasionally retain their source-language inflections even in Late Middle English, a phenomenon which has been categorised as paradigm insertion in Wohlgemuth's (2009) classification of

loan verb accommodation strategies. Our sample contains three attributive loan adjectives which occur with their French-origin -(*e*)*s*-inflections:

(103) *And therfore understoond that bothe he that selleth and he that beyeth <u>thynges</u> **espirituels** been cleped symonyals*

'And understand, therefore, that both he who sells and he who buys spiritual things are called simoniac.' (*The Parson's Tale*)

(104) *But certes , in service , for which men yeven <u>thynges</u> **espirituels** unto hir servauntz , it moot been understonde that the service moot been honest*

'But for sure, in service, for which men give spiritual things to their servants, it must be understood that the service must be honest.' (*The Parson's Tale*)

(105) *in þat chapell syngen <u>prestes</u> **yndyenes***

'In that chapel sing priests of India (lit. Indian priests).' (*Mandeville's Travels*, c1400, PPCME2)

Mustanoja (1960: 277) has indeed documented a tendency for adjectives of French origin, and more specifically in the plural, to occur with -(*e*)*s*-inflections, which is an imitation of French. Also note that the three adjectives in the above examples are post-nominalised in comparison to their nominal heads, namely *thynges* in examples (103) and (104), and *prestes* in example (105). Postposition is generally ascribed to Old French influence in Middle English (see e.g. Jespersen 1949; Mustanoja 1960; Mossé 1991; Wright 2011; Trips 2014), although Attali & Monsonégo (1997) have claimed that postposition in Old French was actually marked. According to Brunner (1963: 284), this is a key feature of some "stereotyped borrowed phrases such as *places delitables* and *lords spirituels*". In most of the cases, the adjective is used postnominally, as in *prestes yndyenes* in example (105), where postnominal use is a direct imitation of French (Brunner 1963: 284).

However, the presence of three cases of paradigm insertion in our subset cannot explain the apparent ease with which some French-origin adjectives are inflected in Middle English. To ascertain whether — and if so, which — factors other than source language may affect inflectional use, we set up an additional mixed-effects logistic regression model. Such a model may reveal any correlations between the different dependent and independent variables. Another advantage of regression models is that lemma frequency can be included in the analysis, since the regression model for verbs in Section 3.3.2.1. has shown that accommodation biases are stronger in lower-frequency items than in higher-frequency items. Therefore, lemma frequency is expected to play a role in loan adjective accommodation as well. The output of the logistic regression analysis is presented in Table 19 below, where inflectional form

(uninflected: 0/inflected: 1) serves as the dependent variable, whereas source language of the adjective, adjectival syntactic position, and lemma frequency are included as independent variables. Potential idiosyncratic authorial, register and textual features were controlled for by including text (i.e. *Mandeville's Travels*, *The Old Testament*, and *The Parson's Tale*) as a random effect in the model. Adjectival lemma was incorporated as a random effect as well. We also ran an interaction between source language and lemma frequency, which can affect the main effects for the variables.

Table 19 displays the coefficient estimates (Est.) for the fixed effects, which are the predicted values of the dependent variable when a certain independent variable. The model also includes standard errors (SE) and the z-value, which shows where the value lies on a normal distribution curve. Last, the model reveals p-values given the z-score, which reveal how likely it is that the sample data would have occurred under the null hypothesis. Apart from the p-values, all numbers in Table 19 are rounded to two digits after comma. The reference levels are English origin for source language, and attributive position for adjectival syntactic position. Frequencies for the individual adjectival lemmas are compared to the mean.

*Table 19: Outcome of mixed-effects logistic regression model for markedness in adjectives (n = 1,887) in* Mandeville's Travels, The Old Testament, *and* The Parson's Tale.

|  | Est. | SE | z-value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | 0.08 | 0.40 | 0.19 | 0.85 |
| Source (French-origin) | -2.77 | 0.76 | -3.63 | 0.000 |
| Frequency | -0.01 | 0.01 | -0.79 | 0.43 |
| Position (predicative) | -1.40 | 0.25 | -5.55 | 2.93e-08 |
| Frequency:Source (French-origin) | 0.15 | 0.10 | 1.53 | 0.13 |

From the output in Table 19 it stands out that adjectival source language, as hypothesised, has a significant effect on inflectional use: the negative coefficient estimate (-2.77) shows that inflection is significantly less prevalent in French-origin than in English-origin adjectives (p < 0.001), which corroborates that French loan adjectives are subject to a markedness bias. The model showcases that lemma frequency (p = 0.43) is not a significant variable in the choice of inflected or uninflected forms. Although the coefficient estimate is negative (-0.01), meaning that low-frequency adjectives are influenced by the markedness bias more strongly than high-frequency adjectives, this effect is not significant. Since the outcome for lemma frequency is not significant, it is not surprising that no significant interaction effect was found between lemma frequency and adjectival source language (p = 0.13). This

means that the preference to use French loan adjectives with uninflected structures does not become even stronger in lower-frequency items as compared to higher-frequency items. This is in contrast with the study on loan verbs, where source language and lemma frequency interacted, meaning that the non-finite bias in French loan verbs becomes even stronger with lower verb frequencies. A last significant variable revealed by the model is adjectival syntactic position (p = 2.93e-08): the negative coefficient estimate (-1.40) shows that predicatives are less likely to be inflected than attributives. This not only corroborates Brunner's (1963: 51) observation, it links the predicative and markedness biases, and can, therefore, also serve as one of the explanations as to why French loan adjectives are subject to a predicative bias.

### 4.3.2.3. Head bias

An additional observation stands out for the attributive adjectives in our sample. French loan adjectives tend to be slightly more frequent with French heads than native English adjectives. In example (106), for instance, both attributive adjective *princypall* and nominal head *cytees* are of French origin.

(106)   *Cypre is right a gode Ile and a fair & a gret and it hath .iiij.* **princypall** <u>*cytees*</u> *within him*

   'Cyprus is a good isle and a beautiful and a great one, and it has four main cities.' (*Mandeville's Travels*)

That French-origin adjectives tend to occur with French-origin heads could have to do with code-switching being restricted between functional heads and their complements, such as attributive adjectives and their associated nouns (Belazi, Rubin & Toribio 1994). Although Belazi, Rubin & Toribio (1994) apply the X-Bar Theory as their framework, this finding translates to other frameworks as well. In Table 20, 43% of the French-origin adjectives versus 38.3% of the English-origin adjectives occur with French-origin heads.

*Table 20: Distribution of French-origin adjectives and English-origin adjectives with French-origin and English-origin heads (Fisher, p = 0.27).*

|  | English-origin heads | French-origin heads |
|---|---|---|
| English-origin adjectives (n = 918) | 566 (61.7%) | 352 (38.3%) |
| French-origin adjectives (n = 172) | 98 (57%) | 74 (43%) |

This observation possibly suggests that French loan adjectives are integrated into Middle English in larger phrasal or syntactic units. However, Trips (2014) has underlined the importance of not considering such French-origin adjectives occurring with same-origin heads as being fixed phrases stored in the lexicon. Also, it becomes clear from Table 20 that the tendency for French-origin adjectives to occur with French-origin heads is not significant (p = 0.27), which may mean that the difference between the proportions is not significant. The findings remain tentative and should, therefore, be treated with caution. The phenomenon of loan adjectives occurring with loan heads will be returned to in Section 5.2., where it will be linked to dual-language activation.

## 4.4. Persistence of loan adjective accommodation biases

Sections 4.2. and 4.3. have provided empirical evidence for the existence of a predicative and markedness bias in both the English-Dutch and French-Late Middle English contact settings. In the present section we will tackle one more question concerning loan word accommodation biases, namely how such biases evolve through time: do they strengthen or weaken, or do they remain stable? In other words, does the stage of adoption in the recipient language affect the strength of the biases?

It is possible that biases wore off with time, as loan words fully integrated into the lexicon, or as authors became more accustomed to accommodating French loans. Therefore, we explore the idea that biases were stronger in Early Middle English (i.e. 1150–1350), when loan words were first adopted, than in Late Middle English (i.e. 1350–1500). Loan adjectives may have been perceived as more foreign in Early Middle English, when borrowing rates were still low, than in Late Middle English. That is because biases in Early Middle English were not as integrated as in Late Middle English, where loan words had already become more conventionalised. To test this hypothesis, we traced the course of the predicative and markedness biases from Early to Late Middle English. The head bias was not looked at since the evidence for adjectives in Late Middle English remains tentative. The reason for focusing solely on the French-Middle English contact setting, and not on the English-Dutch one, is that the data for the former stretches over a considerably longer period of time and can thus draw a more complete picture.

A first critical caveat to this study is that the dataset includes borrowed data, but no baseline of native vocabulary. Instead, the properties of native English adjectives as presented in Section 4.3. are extrapolated to the case study at hand. Second, this study has only been conducted for adjectives, and not for verbs. A third caveat is that many adjectival attestations in this study are attested in the same texts, which means that authorial and/or textual idiosyncratic features were not accounted for. However, due

to the small size of the dataset we did not cap the number of attestations from the same texts. We propose that, despite those caveats, the findings for this case study can help provide insight into how loan word accommodation biases evolve through time. Additionally, "corpus-based research can inform psycholinguistic research in the study of phenomena such as persistence", as seen in Rosemeyer & Schwenter (2019: 199).

## 4.4.1. Data and methodology

### 4.4.1.1. Data extraction

The first step in the data extraction process was to often French-origin adjectives which are attested in both Early and Late Middle English. Since most loan adjectives which entered Early Middle English are still — and even increasingly — used in Late Middle English, the focus was on tracing early French loans. To do so, we ran a query in the online version of the MED, which allowed us to quickly generate a list of loans. The query included adjectives of Anglo-French origin only, since Anglo French was the dominant source-language variety during the first phase of contact, and loans from Anglo French may logically speaking have been borrowed earlier than those from Central French (cf. Section 2.4.3.1.1.). Another way to ensure that the French-origin adjectives in the list occur in the earlier stages of Middle English onwards was to include only search results of texts with a composition date from before 1350, i.e. any attestations before the period defined as M3 (1350–1420) in the HC and PPCME2. Since most French loans entered the language after 1350, we increased the number of adjectives in the list by including all types of genres (i.e. poetry and prose), and adjectives of both high and low frequency in the query. Adjectival attestations which are part of a name (e.g. *Richard le **Demur**) or which are or can be used nominally (e.g. *Austin* 'Augustinian') were not included. This search strategy yielded a list of 52 adjectival lemmas, for which we wrote regular expressions accounting for the spelling variation as reported in the MED (Lewis 1952–2001). An example of the regular expression for French-origin *suspecious* (also written as *suspeciouse, -peccious, -pes(s)ious, susspecious* and *suspicious(e, susspicius, sispicius*; pl. *suspiciouses*) is given in (107):

(107)   \bs[iu]s+p[ei][cs]+io?use?s?\b

Using a specified Perl script, PPCEM2 was then queried for the regular expressions, which generated a random sample containing 35 different adjectival lemmas[52]. A maximum of 5 attestations per period was extracted for each individual adjective. The random sample contained many false positives, such as French-origin noun *power* for

---

[52] The reason for extracting data from the PPCME2 instead of extracting quotations directly from the MED is the potential bias for dictionary makers to collect certain types of data only. As such, the distribution of syntactic position and inflectional endings may be skewed. Although it is not said that this is the case for MED, distributional properties in dictionary material in general should be treated with caution.

French-origin adjective *povre* ('poor'), but the majority of false positives were concentrated around specific lemmas, such as *bene* 'fair, good, pleasant', *fair* 'beautiful', and *vile* 'offensive, useless'. Attestations sporadically contained an erroneous part-of-speech-tag or lemma. After a manual clean-up of the sample, a total of 311 attestations was retained: a subset of 112 attestations for Early Middle English (i.e. 1150–1350) and a subset of 199 attestations for Late Middle English (i.e. 1350–1500). That the number of French loan adjectives in Late Middle English is higher than of Early Middle English is due to the considerably higher number of borrowings in this later phase of contact. Also, Percillier & Trips (2020: 7172) state that "the PPCME2 contains a large data gap in the period M2 (1250–1350)", which "contains only a fourth to a third of the word counts in other periods". According to them, this gap is due to the focus of the corpus builders on prose texts as opposed to poetry. The difference in amounts of data between Early and Late Middle English was accounted for by implementing a baseline of English-origin adjectives in Late Middle English.

### 4.4.1.2. Data annotation

Adjectival attestations, which were all of Anglo-French origin, were manually annotated for their lemma, phase of Middle English, syntactic position and inflectional form. The lemma was the one as identified in the MED. For the phase of Middle English, we distinguished between Early (1150–1350) and Late Middle English (1350–1500), of which the periodisation coincides with the M1–M2 and M3–M4 periods (cf. HC and PPCME2) respectively. Instances of Early and Late Middle English sentences are displayed in (108) and (109) respectively.

(108)   *A gaderyng to-gidres of* **stable** *folk*

'A gathering of stable people' (*The Earliest Complete English Prose Psalter*, c1350, PPCME2)

(109)   *and the trone of the reume of Dauith schal be* **stable** *withouten ende*

'And the throne of David's kingdom shall be stable without end.' (*Purvey's General Prologue to the Bible*, a1450, PPCME2)

For adjectival syntactic position, we reused the categorisation for French loan adjectives in Late Middle English as used in Section 4.3.1.2.: although we distinguished between attributive, predicative, nominal, and adverbial use, only cases of attributive and predicative position were included in data analysis. Annotations for inflectional form, last, distinguished between zero-inflection, schwa-inflection, and forms invariably used with schwa (cf. Section 4.3.1.2.).

### 4.4.1.3. Data analysis

The French-origin adjectives were compared for their adjectival syntactic position and inflectional ending as used in Early and Late Middle English. Those rates were then compared to the English baseline. P-values were retrieved from the Fisher's exact test. We also used R-package "ggplot2" (Wickham 2016) to visualise the data as stacked bar charts, which represent the values for syntactic position and inflectional ending as vertical bars. An important benefit of stacked bar charts — as compared to combined charts — is the ease with which the values for the different categories can be visually compared.

## 4.4.2. Findings

### 4.4.2.1. Predicative bias

Table 21 below reveals how often French loan adjectives in Early and Late Middle English are attested in attributive and predicative syntactic position. Those rates are compared to the Late Middle English baseline of English-origin adjectives (cf. numbers attested in Table 16).

*Table 21: Distribution of French-origin adjectives in attributive and predicative syntactic position in Early and Late Middle English as compared to the English baseline.*

|  | Attributive position | Predicative position |
|---|---|---|
| English-origin adjectives in Late ME (n = 1,171) | 922 (78.7%) | 249 (21.3%) |
| French-origin adjectives in Early ME (n = 77) | 55 (71.4%) | 22 (28.6%) |
| French-origin adjectives in Late ME (n = 160) | 102 (63.3%) | 58 (36.3%) |

It stands out from Table 21 that the French loans in Early Middle English (28.6%) are less common in predicative position than those in Late Middle English (36.3%). However, since this difference is not significant (p = 0.30), the difference between the syntactic position of earlier and later French loan adjectives is not substantial. The dataset is possibly too small for the Fisher's exact test to discern any significant effects. Instead, the main take away from this table is that French loan adjectives are generally more frequent in predicative position than the baseline of native English adjectives. The proportion between syntactic positions in French loan adjectives in Early Middle English and baseline adjectives is not significant (p = 0.15). This is likely due to the small sample size for French-origin adjectives in Early Middle English.

However, the proportion between syntactic positions in French loan adjectives in Late Middle English and native English adjectives is significant (p < 0.001). The proportion of syntactic positions in French loan adjectives from Early and Late Middle English combined compared to baseline adjectives is significant as well (p < 0.001). This implies that French-origin adjectives in both Early and Late Middle English are subject to a predicative bias.

The numbers from the above table are visualised in Figure 9. The rates of French-origin adjectives in attributive and predicative position in Early Middle English (M1–M2 periods) are depicted in the left bar. Those for French-origin adjectives in Late Middle English (M3–M4 periods) are depicted in the right bar. The share of native English adjectives used predicatively is indicated by the horizontal dashed line, which is set at 21%.
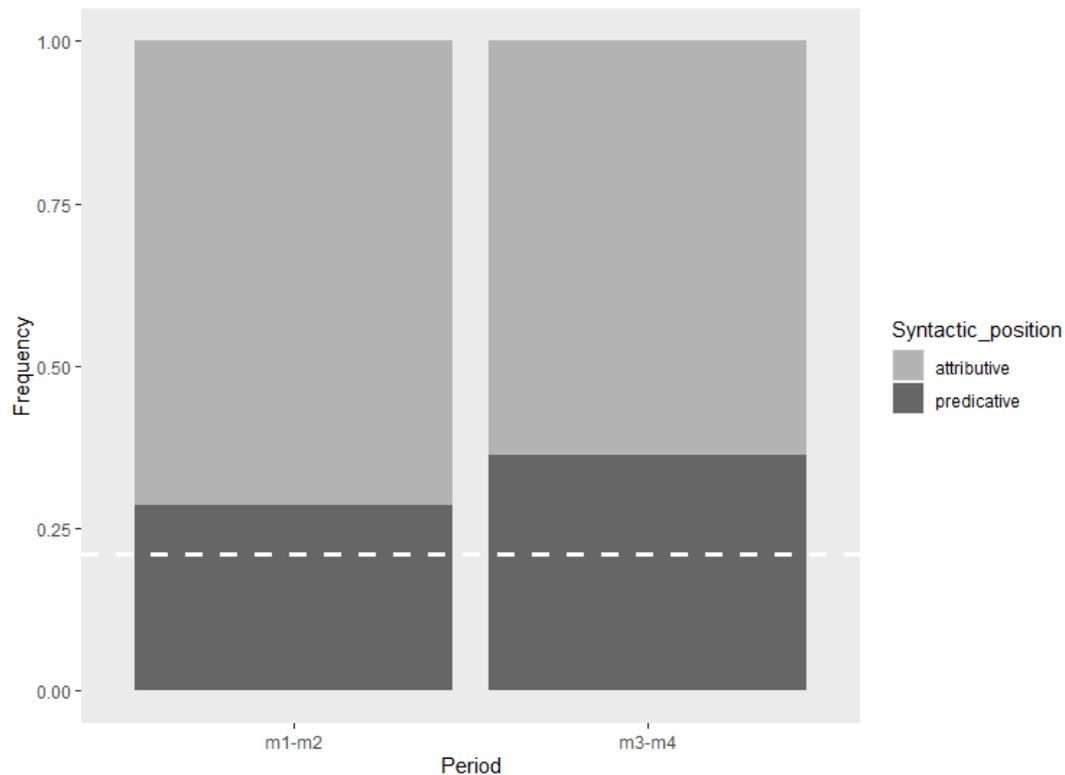


*Figure 9: Distribution of syntactic position for French-origin adjectives in Early and Late Middle English (n = 237) as compared to predicative forms in the English baseline, with horizontal dashed line set at 21%.*

The vertical divisions in the bar chart illustrate that both earlier and later French loan adjectives are subject to a predicative bias. Although we had expected that the predicative bias would be even stronger in earlier French loans than in later loans, this

hypothesis is not borne out. However, with a p-value of 0.15 the proportions are shown not to be significantly different; instead, loan word accommodation biases are shown to be diachronically persistent over a relatively long stretch of time, as Figure 9 shows no indication of the bias weakening in the course of the Middle English period.

### 4.4.2.2. Markedness bias

Table 22 below depicts the rates of French loan adjectives in Early and Late Middle English being used in inflected forms, uninflected forms, or in forms in invariable -*e*. The findings for the French-origin adjectives are compared to the English-origin baseline for Late Middle English (cf. numbers attested in Table 17 and Table 18). Note that this table depicts the rates of inflection for adjectives in both attributive and predicative position combined.

*Table 22: Distribution of French-origin adjectives with inflection, without inflection, and with invariable* -e *in Early and Late Middle English as compared to the English baseline.*

|  | Inflected | Uninflected | Invariable -*e* |
|---|---|---|---|
| English-origin adjectives in Late ME (n = 1,171) | 350 (29.9%) | 458 (39.1%) | 363 (31%) |
| French-origin adjectives in Early ME (n = 112) | 35 (31.3%) | 70 (62.5%) | 7 (6.3%) |
| French-origin adjectives in Late ME (n = 199) | 97 (48.7%) | 92 (46.2%) | 10 (5%) |

Table 22 shows that earlier French loan adjectives (62.5%) are more common in uninflected forms than later French loan adjectives (46.2%), and this difference is strongly significant (p = 0.003[53]). This means that the markedness bias in earlier loans is even stronger than in later loans. As to a comparison between French-origin and English-origin adjectives, the earlier French loans are more frequent in uninflected forms than the English baseline (39.1%), but this proportion is only marginally significant (p = 0.06). The later French loans, too, are more frequent in uninflected forms than the English baseline, but the difference in proportions is only borderline significant (p = 0.05). When the proportions of French-origin adjectives in both periods are compared to those of English-origin adjectives, the difference is not significant (p = 0.68). Therefore, we can conclude that there are no significant

---

[53] The p-values in this section depict the distribution of inflected and uninflected forms. Forms in invariable -*e* are not taken into account in the Fisher's exact test.

differences between the markedness bias in Early and Late Middle English, and that the bias is relatively stable.

The numbers shown in Table 22 are also visualised in the stacked bar charts in Figure 10, where the rate of inflected forms in the English baseline (43%) is represented by the horizontal dashed line. The distribution of inflected and uninflected forms in French loans in Early Middle English is depicted in the left bar (M1–M2 periods); the distribution in French loans in Late Middle English is depicted in the right bar (M3–M4 periods). Note that the darkest bar represents inflected forms, which are avoided, and the bar lightest in colour represents uninflected forms.
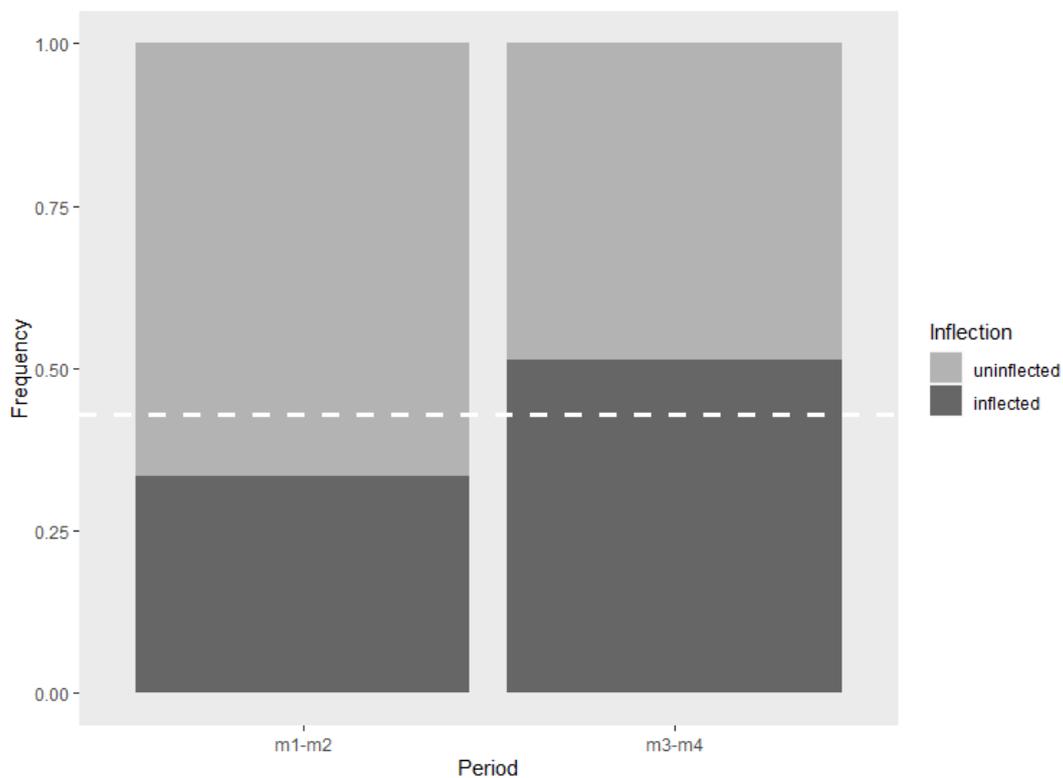


*Figure 10: Distribution of inflection for French-origin adjectives in Early and Late Middle English (n = 311) as compared to inflected forms in the English baseline, with horizontal dashed line set at 43%.*

The vertical divisions for inflectional ending confirm that earlier loans have considerably fewer inflections than later loans, which may relate to the lower levels of conventionalisation of earlier loans. To conclude, this section has shown that the use of inflection increases over time, namely when loan adjectives become more integrated and conventionalised in their recipient language.

## 4.5. Discussion

The present chapter has investigated the phenomenon of loan word accommodation biases in English loans in Present-day Dutch and French loans in Late Middle English. Whereas the focus of Chapter 3. was on verbs, the focus of this chapter was on the category of adjectives. The reason for investigating adjectives was the formal resemblance between non-finite verbs and predicative adjectives, which both typically occur in combination with an auxiliary or copular verb carrying the grammatical information. Also, both categories are typically used in clause-final position. This is in stark contrast with finite verbs and attributive adjectives, which tend to be used clause-medially and which are not combined with an auxiliary or copular verb. Our findings corroborate that the loan word accommodation biases in English loans in Present-day Dutch and French loans in Late Middle English are not limited to the category of verbs, but extend to adjectives. Similar to verbs, loan adjectives in both language contact settings are subject to a syntactic bias (*predicative bias*) and a morphological bias (*markedness bias*). The predicative bias, first, refers to language users exhibiting a higher preference to use loan adjectives in predicative syntactic position than in attributive syntactic position. Evidence for this bias can be found in both contact settings under investigation. That predicative syntactic position is preferred over attributive position ties in with the "developmental lag" (Ninio 2004: 258) of attributive adjectives in first-language acquisition, which shows that attributive position comes with an elevated morphosyntactic complexity compared to predicative position. This will be discussed in detail in Section 5.3.1. Under the markedness bias, second, loan adjectives tend to be preferred in uninflected forms or less explicitly marked categories as compared to inflected forms or more explicitly marked categories. In both contact settings, there is some evidence for loan adjectives to be favoured without inflection. For Late Middle English, the data additionally reveal occasional retention of French inflection in French loan adjectives, a phenomenon which has been called paradigm insertion in Wohlgemuth's (2009) classification of loan verb accommodation strategies. The data for Late Middle English also revealed tentative evidence for a third bias (*head bias*), namely for French-origin adjectives — as compared to English-origin adjectives — to be more common with French-origin nominal heads. Although this suggests that French loan adjectives may have entered the Middle English language in larger syntactic units, the evidence for the head bias is not significant and should thus be interpreted with caution.

In both language contact settings, the nature of the accommodation biases for loan adjectives is similar to that for loan verbs. That is because loans of both parts of speech are subject to a syntactic and a morphological bias, with the evidence for the syntactic bias consistently ranking stronger than for the morphological bias. First, the

differences between native and loan adjectives are larger for the predicative bias than for the markedness bias. Second, when the predicative and markedness biases enter into conflict — for instance in the case of a predicative, inflected adjective — predicative position is generally preferred over attributive position, even if that implies that the loan adjective is inflected. In sum, the predicative bias is generally more pervasive than the markedness bias. However, notice that predicative adjectives in Dutch are never inflected (Treffers-Daller 1994: 144), and those in Late Middle English less often than attributive adjectives (Brunner 1963: 51). As a result, the preference for predicative syntactic position and for uninflected forms cannot reasonably be detached, and the two biases are deeply intertwined. Although the evidence for the markedness bias is less strong than for the predicative bias, the findings for loan adjectives add to the assertion that even direct insertion (Wohlgemuth 2009), the most frequent loan word accommodation strategy, is not itself free from constraints, and that borrowing is restricted by inflectional endings. By means of loan word accommodation biases, language users may subconsciously try to lower the increased processing cost coming with the borrowing of loan adjectives (cf. Section 5.3.4.). This adds to previous research by, among others, Harris & Campbell (1995) and Schultze-Berndt (2017: 265), as can be read in Chapter 3.

Section 4.4. contained a smaller case study which documented the trajectory of accommodation biases in loan adjectives through time, assessing whether they strengthen, weaken, or remain stable. This hypothesis was tested for the French-Middle English contact setting only since its history of language contact has been completed, as opposed to the English-Dutch contact setting, where contact is still ongoing. We hypothesised that the predicative and markedness biases would be even stronger in more recent (Early Middle English) than in more established loans (Late Middle English), since the increased processing cost coming with loan words (cf. Section 5.3.4.) may be even higher in more earlier than in later loans. The findings for the predicative bias revealed some differences between the rates of earlier and later loans, but they are not significant. Also, French loan adjectives from both periods of Middle English are — as expected — more common in predicative syntactic position than native English adjectives. This indicates that the predicative bias stays remarkably persistent over a longer stretch of time. As to the markedness bias, earlier French loan adjectives are significantly more common in uninflected forms than later French loans, revealing that the bias in earlier loans is even stronger. That is because earlier loan words limit processing more than later loan words, and the processing cost of earlier loans decreases when they become (or are considered) increasingly native-like over time. This argument will be further developed in Section 5.3.4. It may be instructive to extend the present corpus study beyond the Middle English period into the Early Modern English period to learn how long accommodation biases

persist. The expectation is that biases in Early Modern English are weaker than in Middle English, or that they even have disappeared altogether.

It should be noted that loan word accommodation biases in this thesis are tested in two specific language contact settings with similar language pairs. Although caution is in place when generalising the findings, some researchers in the past have already hinted at accommodation biases being at hand in some non-Indo-European languages in contact as well. More details on this caveat can be found in Section 3.4.

This chapter deepens our general understanding of loan word accommodation — as well as accommodation biases — in different parts of speech. The existence (and persistence) of loan word accommodation biases in verbs and adjectives evokes a range of new research questions. One of the main questions which arises is which theoretical explanations can account for the existence of biases in loan words, which will be the topic of the next chapter.

# Chapter 5

# Causes of accommodation biases[54]

## 5.1. Introduction

The previous two chapters have provided extensive evidence for the existence of loan word accommodation biases in verbs and adjectives. More specifically, biases are attested in English loans in Present-day Dutch and French loans in Late Middle English. For loan verbs, there is a tendency towards non-finite forms (*non-finite bias*) and uninflected forms (*markedness bias*) compared to native verbs; for loan adjectives, there is a tendency towards predicative forms (*predicative bias*) and uninflected forms (*markedness bias*) compared to native adjectives. To put it another way, both parts of speech are subject to a syntactic and a morphological bias, with the evidence for the syntactic bias being consistently stronger. An additional (and smaller) case study tracing the evolution of the predicative and markedness biases in French loan adjectives through Early and Late Middle English has revealed that accommodation biases are remarkably persistent over longer stretches of time.

Chapters 3. and 4. have already touched upon two factors which do NOT cause accommodation biases. First, adding inflectional morphology to a loan stem is an obstacle to loan word integration, but it is not the main one. This is because both loan verbs and adjectives are affected by a markedness bias, which indicates that the addition of inflections to loan stems is an obstacle to loan word integration (see e.g. Harris & Campbell 1995: 135; Sijs 2005: 56–57). This finding refuted Wohlgemuth's (2009: 291) argumentation that — since direct insertion is the most frequently used loan word accommodation strategy cross-linguistically — loan word integration is not constrained by inflectional morphology. However, for both parts of speech we have also shown that the markedness bias is, in fact, only of secondary importance, since the major biases (i.e. non-finite and predicative biases) are of a syntactic nature. For example, the preferred usage categories for loan verbs in Dutch and Middle English include past participles (e.g. *ge-check-t* in Dutch) and infinitives (e.g. *check-en* in Dutch), both forms which carry inflections. This means that the

---

[54] The argumentation in this chapter is based on parts of a manuscript submitted for publication (De Smet & Shaw subm.).

syntactic bias dominates over the morphological bias. Therefore, inflection cannot be held accountable for the existence and persistence of loan word accommodation biases. A second factor which does not seem to trigger the existence of biases is the properties and distributional tendencies of the source language, which may differ from recipient-language tendencies. In Present-day English, for instance, the distribution of non-finite and finite verb forms is similar to the distribution in Present-day Dutch. Both languages have similar distributional tendencies for attributive and predicative syntactic position in adjectives as well. The same can be said for French and Middle English. The question then presents itself: which factor or factors CAN account for the existence of accommodation biases in two language contact settings with similar language pairs?

In this chapter we will provide the theoretical background and explanations which could aid in demystifying the presence of loan word accommodation biases. Section 5.2. will explain how the processing cost of loan words, which will be defined below, is increased by dual-language activation in bilingual speakers. Consequently, the presence of loans complicates language selection monitoring, and potentially slows down language users in their processing of loans. We will argue in Section 5.3. that accommodation biases can act as a facilitative strategy to reduce that increased processing cost of loans. That is, loans are used in contexts which store less grammatical information, such as non-finites and predicatives, where auxiliaries hold most of the information. The reasoning developed in Sections 5.2. and 5.3. will allow us to position accommodation biases with respect to the continuing debate on the difference between lexical borrowing and code-switching, and their interdependent relationship (Section 5.4.). Code-switching and borrowing, which have been defined in the state of the art (Section 2.2.), have features in common, since language users keep associating even well-integrated borrowings with their source language. We will argue that loan word accommodation biases, which share features with both phenomena, characterise the gradual transition from less-integrated code-switch to highly conventionalised borrowing. Section 5.5. will touch upon some possible alternative factors which may play a role in the emergence of accommodation biases. A general discussion and conclusion of the theoretical argumentation will be presented in Section 5.6.

## 5.2. Loans and their processing cost

In their research on bilingualism, de Groot & Starreveld (2015: 390) have provided extensive evidence for "word recognition in bilinguals [being] language non-selective". This means that lexicons from both languages are activated simultaneously in both word recognition and production (see e.g. Levelt et al. 1991; Grainger & Dijkstra 1992; Bijeljac-Babic, Biardeau & Grainger 1997; van Heuven,

Dijkstra & Grainger 1998; Marian & Spivey 2003; Weber & Cutler 2004; Blumenfeld & Marian 2007). The process of "language non-selectivity" (Groot & Starreveld 2015: 414) will be referred to as *dual-language activation* throughout this thesis. As briefly touched upon in Chapters 3. and 4., the morphosyntactic integration of loan words seems to be associated with an added processing cost (also called processing load or production cost). The added cost is somewhat surprising, since direct insertion is the loan word accommodation strategy requiring the least integrational effort (cf. Wohlgemuth 2009: 135). That is because the verb is immediately fully functional and does not need additional modification. However, we will now argue that the elevated cost associated with loan words could arise because speakers are aware that loan words belong to a different language. This may increase dual-language activation. The term 'increased' (as opposed to normal) dual-language activation in this thesis is adopted to refer to higher levels of dual-language activation than measured on average. It does not mean that dual-language activation occurs where it is not expected or required, since it is basically present in all bilingual phenomena (cf. Dijkstra 2005; infra). Another concept which should be properly delineated is *processing cost*, which in this thesis is defined as the cognitive cost emerging (i) while producing a word or string of words oneself, or (ii) while processing a word or string of words that has been uttered by one's interlocutor (i.e. a person or a text). Increased processing cost of a word or string of words is linked to increased levels of difficulty of that word or phrase.

Dual-language activation is essentially present in all bilingual phenomena, since both languages are always activated in bilingual speakers' minds (e.g. Dijkstra 2005). However, this practice can be boosted by the emergence of a cognate[55], or "a word that not only shares meaning with its translation in the other language but also all, or a large part, of its form" (Groot & Starreveld 2015: 390). An example of a cognate is *table*, which refers to the same piece of furniture in both English (*a table – tables*) and French (*une table – les tables*), and thus is an English-French cognate for bilingual speakers of those two languages (Groot & Starreveld 2015: 390). Evidence for cognates boosting dual-language activation has been provided by, among others, Li & Gollan (2018: 933), who have shown in their experiment on Mandarin-English bilinguals that — when reading out loud — the presence of cognates considerably "increases dual-language activation at a stage of processing in which translation equivalents compete for selection". According to their findings, dual-language activation is stronger in example (110), which contains a Mandarin-English cognate

---

[55] Some authors would call cognates with identical meanings 'true cognates' (e.g. De Cecco 2017: 270). Also note that this chapter does not expand on the phonological features of cognates.

(*coffee* versus *ka1 fei1*), than in example (111), which contains a non-cognate (*tea* versus *cha2*[56]). Both examples are cited from Li & Gollan (2018: 934).

(110)   *The woman behind the table made an unknown hand gesture over a cup of hot* 咖啡 *(i.e. ka1 fei1 'coffee').*

(111)   *The woman behind the table made an unknown hand gesture over a cup of hot* 茶 *(i.e. cha2 'tea').*

On a side note, one should bear in mind that contact between Mandarin and English differs rather strongly from the language pairs investigated in this thesis, since Mandarin and English are unrelated languages, whereas the other language pairs are (Indo-)European languages. This may also mean that most or all cognates in the Mandarin-English contact setting are loans, whereas in the contact settings under investigation many cognates are cognates without being loans. This is due to the difference in relatedness between the languages involved in contact.

Various priming studies have shown that cognates can speed up lexical retrieval (e.g. Costa, Caramazza & Sebastian-Galles 2000; Hoshino & Kroll 2008). While this indicates that dual-language activation facilitates processing, increased dual-language activation due to cognates may also considerably increase bilingual speakers' effort to monitor language selection (Li & Gollan 2018: 933). Therefore, it may be linked to higher levels of processing limitations.

For bilingual speakers, many loan items are cognates[57]. Since cognates can boost dual-language activation and complicate language selection, loan items can — like cognates — be expected to trigger increased activation of the source language. For instance, in Li & Gollan's (2018: 924) experiment bilingual speakers produced significantly more intrusion errors, or "the selection of the right word but in the wrong language", when reading sentences containing cognates than when reading sentences without cognates. At the same time, intrusion errors coming with cognates were also identified and corrected significantly more quickly than with non-cognates. Another reason why cognates may boost source-language activation is the "triggering hypothesis" (e.g. Clyne 1980), according to which the presence of cognates facilitates subsequent code-switching in spoken language. This is illustrated in example (112), which shows an utterance by a Dutch-English bilingual speaker and which is reproduced from Broersma (2009: 456). The utterance in (112) follows an excerpt containing exclusively Dutch text. Note that all English-origin material is printed in

---

[56] However, *cha* (also *chah*, *chai*, *char*, *chi*) is actually attested to mean 'tea' in the phrase 'a cuppa **cha**' used in informal English (Green's Dictionary of Slang 2022).
[57] Notice that the activation literature does not differentiate between cognates and loans (e.g. Groot & Starreveld 2015; Li & Gollan 2018).

bold, and that all code-switches are underlined. The reason why Dutch *en* and English *and* in (112) are grouped together is that both words sound similar, and from this excerpt Broersma (2009) could not reliably infer whether the speaker uttered the Dutch or the English conjunction.

(112)  *[En /* **And***] **apparently** deze mevrouw was* **called** *R., net* **like me***.

'And apparently this lady was called R., just like me.' (Broersma 2009: 456)

Following Broersma (2009: 456), the behaviour of the code-switches in example (112) can be described as follows: the first code-switch is *apparently*, which switches from Dutch to English[58]. *Apparently* is followed by cognate *deze*, which switches from English back to Dutch. *Was* is considered a cognate, after which the speaker switches to English (*called*) and utters a proper noun (*R.*), which is the name of the speaker. Then the speaker code-switches to Dutch (*net*), and then back to English (*like*). Note that Broersma (2009: 456) refers to cognates as *trigger words*, which fits the triggering theory framework. Trigger words, or cognates, alternate with code-switches following a regular pattern. That is, "all the words that are not trigger words are codeswitched, and after each trigger word follows a codeswitch" (Broersma 2009: 456).

In our own data as well, we have found tentative evidence of increased dual-language activation through loan words. By means of illustration, in Section 4.3.2.3. we have described how French-origin adjectives — as compared to English-origin adjectives — tend to be more frequent with French-origin nominal heads. One of the examples to illustrate this head bias is reproduced in (113), where attributive adjective *princypall* 'main' and nominal head *cytees* 'cities' are both borrowed from French and are used together in one noun phrase:

(113)  *Cypre is right a gode Ile and a fair & a gret and it hath .iiij.* **princypall** *cytees within him*

'Cyprus is a good isle and a beautiful and a great one, and it has four main cities.' (*Mandeville's Travels*, c1400, PPCME2)

In some cases, borrowings keep carrying source-language inflections even in their recipient language, which is a case of paradigm insertion (cf. Wohlgemuth 2009). An example is provided in (114), where French loan word *yndyenes* 'Indian', which modifies French-origin head *prestes* 'priests', carries its original French -*s*-plural in Middle English.

---

[58] This means that Broersma (2009: 456) suggests that Dutch *en* is used, and not English *and*.

(114)   *in þat chapell syngen <u>prestes</u>* **yndyenes**

'In that chapel sing priests of India (lit. Indian priests).' (*Mandeville's Travels*, c1400, PPCME2)

The use of a French-origin noun or adjective in Middle English may, as it were, trigger the use of another French-origin noun or adjective rather than an English-origin one. This ties in with findings from code-switching research showing that switches mainly occur between a determiner and noun-adjective combination, as opposed to between a noun and adjective within the same noun phrase (Parafita Couto & Gullberg 2019). Those observations are based on code-switches between Welsh and English, Spanish and English, and Papiamento and Dutch.

This section has shown that cognates complicate language selection and boost dual-language activation. Therefore, loans are expected to trigger increased source-language activation. Since loans tend to be persistently associated to their source language, language users may pay an increased processing cost when integrating loan words in their recipient language.

## 5.3. Accommodation biases as facilitative strategies for loans

Accommodation biases are reminiscent of well-documented usage tendencies observed in other types of linguistic research which can be plausibly linked to processing limitations. The next paragraphs discuss examples from research on first-language acquisition (Section 5.3.1.), lower-frequency items in both first and second languages (Section 5.3.2.), and intrasentential (i.e. within the sentence) code-switches (Section 5.3.3.). This section will draw parallels between these phenomena and loan words, and it will link the increased processing cost associated with loan words to the presence of accommodation biases, which contribute in reducing that cost.

### 5.3.1. Processing limitations in first-language acquisition

One of the most common goals of first-language acquisition research is to predict what children[59] will learn, and in what order. For verbs, children growing up in English — but also in many other languages, such as French, Dutch, German, Swedish, and Norwegian — typically acquire compound or lexical basic forms, such as infinitives, before they acquire finite forms (Kampen & Wijnen 2000; also see Section 3.3.2.). This is illustrated for Dutch in example (115) below, cited from Kampen & Wijnen (2000: 249). The example is uttered by a two-year-old Dutch-speaking child.

---

[59] The focus in this thesis is on neurotypical children. We do not focus on neuroatypical children, or subjects who — due to various circumstances — only start to learn to speak their first language as adolescents or adults.

(115)  *aap*          *banaan*          **eten**

       'monkey        banana            eat.INF'

In example (115), the child uses a non-finite verb form in a construction where adults would likely have used a finite verb form instead, or where they would have used a finite form additionally to the non-finite form. That is, an adult would most likely have produced a simple present (*De aap* **eet** *een banaan* 'The monkey eats a banana') or a present progressive form (*De aap is een banaan aan het* **eten** 'The monkey is eating a banana'). In the case of the simple present construction, the finite form of the lexical verb simply replaces the non-finite infinitival form used by the child in (115), hence *eet* instead of *eten*. This is different for the present progressive, which is formed by means of an inflected form of copula *zijn* ('be') combined with *aan het* ('at the') and the infinitive of the lexical verb. In that case, a finite and non-finite form are combined into one construction. Note that what the child actually wanted to convey cannot reliably be proven, but what is certain is that the child in example (115) does not use a finite form whereas an adult would have used one. Such a preference for non-finite forms has also been documented by Wexler (1994: 308), who claims that young children tend to go through an "'optional infinitive' stage". According to Wijnen (2000), this is the third stage of the time span in which finiteness markers are acquired; however, in this thesis we will not address the first and second stages, which are the *infinitive stage* and the *lexical-finite stage* respectively. In the optional infinitive stage, which lasts until roughly two and a half years, "children produce both finite and non-finite forms in matrix sentences" (Wexler 1994: 308) in positions where adults use finite forms only. This means that, in the child's syntax, finiteness marking is optional, whereas in the adult's syntax it is compulsory. The optional infinitive stage has been linked to children's processing capacity being more limited than that of adults (cf. Lasser 2002: 780–781). According to Phillips (2010), for instance, children experience difficulties in accessing and applying morphological knowledge, which explains why they tend to use infinitives instead of finite forms, which need to be inflected. Optional infinitives have — apart from in child language — also been observed in foreigner talk (cf. Muysken 2013) and in motherese (cf. Rice, Noll & Grimm 1997), as well as in speakers with language impairment (cf. Rice & Wexler 1996; Brunger & Henry 1998). The tendency towards infinitives increases as the processing cost of forms increases. This finding is in line with the findings from Chapter 3., since loan verbs are avoided in finite forms, which may reflect the increased processing capacity needed to produce finite verb forms.

Another example from first-language acquisition research is the acquisition of adjectives, which is less-documented than that of verbs. Research has revealed some evidence of a "developmental lag" (Ninio 2004: 258) of attributive adjectives. For

example, in her research on the use of attributive adjectives by young children, Nelson (1976) has observed that children between two and two and a half years old use adjectives mainly predicatively. Over time children start using adjectives more in attributive syntactic position as well. Similar findings have been presented by Tomasello (1992), who has shown that his two-year-old daughter used adjectives exclusively predicatively at the two-word stage. Based on both a literature review and his own research, Ninio (2004: 259) has likewise found several indications of a "developmental delay of attributive adjective-noun combinations in production" in Hebrew, where adjectives behave similarly to English and French adjectives (Ninio 2004: 261). This finding points to the higher levels of difficulty in acquiring attributive adjectival structures. However, Ninio (2004: 259) has added that "it is very difficult to prove this conclusively" since children's utterances containing adjectives cannot always reliably be classified as predicative or attributive, even more so in the two-word stage. In the case of the utterance *small book*, for example, one interpretation is that *small* is used attributively compared to *book*, and that the adjective is used in its usual position in English (i.e. at the left of its head). However, another interpretation is that *small* is used predicatively, and that copula *be* is elided. In that case, the adjective is not used in its usual position, namely at the right of its head, as in *The book (is) small*. Following Braine & Bowerman (1976: 76), the second interpretation — where *small* is analysed as a predicative adjective — is most likely, since "children's possessive and attributive combinations mostly occur in isolation and do not seem to serve [the] function [of indicating the argument of a predicate]". Thus, in the case of *small book*, it would be less likely that the child comments on what books exactly are "involved in some activity or state of affairs" than that the child predicates the size of the book. Despite the evidence of the attributive lag remaining tentative overall, Ninio (2004) has concluded that it reflects the elevated morphosyntactic complexity of attributive syntactic position compared to predicative syntactic position in the languages studied. Based on logicians (e.g. Montague 1974), Ninio (2004: 256) has attributed this delay to the process in which attributive adjectives are "interpreted relative to the noun" not being fully automatized yet in children. This finding is in line with the predicative bias presented in Chapter 4. That attributive adjectives are avoided in loans may indicate that they involve an increased morphosyntactic complexity compared to predicative adjectives, as found in first-language acquisition research. However, note that most first-language acquisition research builds on data from English and other widely-spoken European languages, and that the above claims only refer to those languages, and not to cross-linguistic tendencies. In fact, many languages do not have a clearly delineated category of adjectives. Then again, this thesis focuses on European languages and can, therefore, follow the claims made by Ninio (2004) and proponents.

## 5.3.2. Processing limitations in low-frequency items

Processing limitations can also be observed in lower-frequency items in both first and second languages.

For second languages specifically, Chapters 3. and 4. have showcased how the biases of loan words towards non-finite and predicative structures are attested in all verbs, hence also in lower-frequency items. In Middle English verbs, source language and lemma frequency interact, meaning that the biases for French loan verbs become even stronger in lower-frequency verbs.

Research from the cognitive linguistics framework has demonstrated that lower-frequency items are generally less entrenched than their higher-frequency counterparts (e.g. Langacker 1987; Bybee 2010; Schmid 2016). The fact that frequency of use impacts the entrenchment of a structure becomes clear from, among others, Langacker's (1987: 59) seminal definition of entrenchment:

> Every use of a structure has a positive impact on its degree of entrenchment, whereas extended periods of disuse have a negative impact. With repeated use, a novel structure becomes progressively entrenched, to the point of becoming a unit; moreover, units are variably entrenched depending on the frequency of their occurrence.

He has stated that frequency of use and the degree of entrenchment are directly proportional. That correlation between frequency of use and entrenchment of usage patterns correlate has been acknowledged by various researchers (see e.g. Diessel 2007; Stammers & Deuchar 2012; Schmid 2013; Divjak & Caldwell-Harris 2015; Schmid & Mantlik 2015; Rosemeyer & Schwenter 2019). A cognitive explanation of this direct proportionality is provided below.

> [T]he productive or receptive processing of a given usage event will leave a memory trace of the neuronal and cognitive patterns of activation required for processing it. If the same or a similar usage event recurs several times, the pattern of activation will be strengthened and/or the commonalities of these usage events will be 'stored' in the form of a second-order 'representation', i.e. a schema. (Schmid & Mantlik 2015: 587)

In other words, since lower-frequency items leave fewer "memory trace[s] of the neuronal and cognitive patterns of activation" (Schmid & Mantlik 2015: 587) than higher-frequency items, the activation process will not be strengthened as much, which leads to an increased processing cost for lower-frequency items. That is, lower-frequency items can be linked to greater processing difficulties than their higher-frequency counterparts. Neurocognitive research has corroborated the finding that higher-frequency items have stronger associative links than lower-frequency

items, and that they are thus easier to access from the lexicon than lower-frequency items (e.g. Berglund-Barraza et al. 2019: 2). More concretely, "word frequency effects can be seen in listeners' quicker reaction times and greater accuracy in accessing high-frequency as compared to low-frequency words from their lexicon, in recalling high as compared to low-frequency word [sic] in serial recall tasks, and recognizing high- as compared to low-frequency words in lexical decision tasks" (Berglund-Barraza et al. 2019: 2). This is not only the case for reading, but also for psycholinguistic tasks, such as lexical decision-making, picture naming, and auditory word comprehension (for references, see Desai, Choi & Henderson 2020: 583). Therefore, the processing cost of lower-frequency items is observed to be higher than that of higher-frequency items (cf. Berglund-Barraza et al. 2019; Desai, Choi & Henderson 2020).

### 5.3.3. Processing limitations in code-switching

A last example of a tendency linked to processing limitations can be found in synchronic code-switching research, where the term 'matrix language' (also called default language) is often used to refer to "the language with the most critical grammatical contributions to the bilingual clause" (Myers-Scotton & Jake 2014: 3). Establishing a matrix language for an *embedded language* (also called non-default language) is necessary to define the directionality of the switch, which means determining whether the language user switches from language X to Y, or from language Y to X (e.g. Raichlin, Walters & Altman 2018: 630). The role of matrix and embedded languages will be further discussed in Section 5.4.1.

Many code-switching researchers have shown through longer response times that language users pay a switch (also called switching) cost when code-switching (e.g. Meuter & Allport 1999; Verhoef, Roelofs & Chwilla 2009; Bobb & Wodniecka 2013; Deibel 2020; for an overview, see Li, Ferreira & Gollan 2022)[60]. This cost has been linked to, among others, the grammatical distance of the languages in the language pair, and the cost has turned out to be lower when the languages have no shared lexical roots (Deibel 2020). However, the extent and applicability of the switch cost are rather controversial: "findings of bilingual processing costs are increasingly acknowledged to be contingent on study participants, experimental design and language mixing type" (Torres Cacoullos 2020: 2). The focus in this section is on the morphosyntactic limitations associated with switches. Concrete examples come from synchronic research on single-word code-switches, specifically in cases where the inserted item preserves source-language inflections. Structurally

---

[60] However, Gullberg (2022) has shown that code-switching is not necessarily costly. This is particularly true for those cases where language users have the choice whether to switch or not. In a similar vein, Adamou & Shen (2019) have shown that code-switching is not costly when it is frequent.

speaking, we have noticed that the loan word accommodation biases found in Chapters 3. and 4. strongly resemble some of the probabilistic morphosyntactic constraints on intrasentential (i.e. within a sentence) code-switching. For verbs, for instance, there exists convincing evidence that inserted other-language verbs are more common in non-finite than in finite verb forms (Myers-Scotton & Jake 2014; also see Keller 2020: 66). Myers-Scotton & Jake's (2014) argument is based on a wide range of languages from different language families, such as Acholi, English, Ewe, Hungarian, Moroccan Arabic, Swahili, and Wolof. This cross-linguistic tendency towards non-finite forms strongly reflects the non-finite bias documented in Chapter 3. and is explained as a facilitative strategy for language users to reduce the morphosyntactic processing cost of inserted items (Myers-Scotton & Jake 2014: 8).

For adjectives as well, there exists evidence for tendencies linked to processing difficulties, yet it is more tentative. Myers-Scotton (2002: 132), for example, has observed that code-switching corpora generally contain a low rate of attributive adjectives, or "Embedded Language adjectives modifying Matrix Language" [61]. In fact, corpora generally contain few adjectives in adjective-noun (i.e. attributive) combinations (Parafita Couto & Gullberg 2019: 702), which cannot be excluded as a reason for code-switches being less common in attributive adjectives than in predicative adjectives. However, this underrepresentation of attributive attestations in code-switches may also reflect the structural complexity of attributive adjectives, which relates to "congruence problems at all levels" (Myers-Scotton 2002: 132). Such congruence problems have also been identified by Deuchar (2005: 257) in her research on Welsh and English. She has distinguished between *paradigmatic congruence*, which refers to the "similarity or equivalence between the grammatical categories of two languages", and *syntagmatic congruence*, which refers to "similarity of word order". Whereas she has found that English adjectives in Welsh are inserted more frequently in predicative position (71%), adjectives are inserted less frequently in attributive position (29%), and she has attributed this tendency to a difference in congruence of predicative and attributive adjectives in Welsh and English. More concretely, in predicative position there is both paradigmatic and syntagmatic congruence between the two languages, since the grammatical categories as well as the word orders are congruent. However, in attributive position there is less syntagmatic congruence, since attributives in Welsh are used mainly postnominally, whereas in English they are used mainly prenominally (Deuchar 2005: 263). These findings seem to refine Poplack's (1980: 586) seminal *equivalence constraint*, which is part of her Constraints model:

---

[61] However, the number of available code-switching corpora is limited.

> Code-switches will tend to occur at points in discourse where juxtaposition of L1 and L2 elements does not violate a syntactic rule of either language, i.e. at points around which the surface structures of the two languages map onto each other.

The equivalence constraint essentially states that "a switch is inhibited from occurring within a constituent generated by a rule from one language which is not shared by the other" (Poplack 1980: 586). In other words, switching is only possible when orders overlap.

Following Treffers-Daller's (1994: 144) study on code-switches from Brussels French into Brussels Dutch, differences in the rates of inserted predicative and attributive adjectives can also be linked to morphological constraints. In Dutch, for instance, adjectives in attributive position can be inflected, whereas in predicative position they cannot be inflected (cf. Section 4.2.1.2.). As a result, "predicative position is more readily accessible for French adjectives, because they do not receive Dutch inflection there" (Treffers-Daller 1994: 144). This means that it would be easier for a language user to insert a French-origin adjective in Dutch in predicative than in attributive position. This is in accordance with our own findings, namely the presence of a predicative bias in loan words (cf. Chapter 4.).

## 5.3.4. Processing limitations in loan words

Like the examples from other domains of linguistic research, and as shown in Section 5.2., loan words may put a burden on processing. In this section we suggest that accommodation biases may serve as a tool to facilitate the morphosyntactic integration of loan words.

As described in Section 5.1., the major accommodation biases found in this thesis (i.e. non-finite and predicative biases) are of a syntactic nature. Structurally speaking, language users seem to prefer periphrastic strategies to operationalise loan words in their recipient language[62]. This can be seen in, for instance, the non-finite bias found in loan verbs in Present-day Dutch and Middle English. Whereas finite verbs are portmanteau forms which have to be inflected for tense, person, and number, non-finite verbs are used with an additional auxiliary, and they only occasionally have a subject. In the case of perfect construction *he has* **helped**, for example, auxiliary verb *has* carries information on tense (present), person (third), and number (singular), whereas *helped* is used as an invariable past participle. It is not a coincidence that the non-finite form, which carries less grammatical information, is the preferred form for loan verbs. A similar preference for periphrastic structures is observed for adjectives, where predicative position is favoured significantly over attributive position. Whereas

---

[62] The use of non-finite verbs and predicative adjectives as periphrastic structures has also been discussed in Chapter 2.

attributive adjectives in Dutch and Middle English can be inflected for person, gender, and number, predicative adjectives in Dutch are categorically not inflected, and in Middle English only sporadically. Dutch *sterk* ('strong'), for instance, may carry inflections in attributive use, as in *een* **sterk<u>e</u>** *vrouw* ('a strong woman'); however, in the case of predicatives, adjectives are combined with a copula, as in *die vrouw <u>is</u>* **sterk** ('that woman is strong'). In both predicative adjectives and non-finite verbs, it is the auxiliary, modal or copular verb which carries the grammatical information, which results in less information being coded directly on the loan word. Simply put, using loan words in periphrastic structures equals using loan words in morphosyntactically less complex structures. As such, using periphrastic forms may lower the processing cost of morphosyntactic paradigms, by avoiding the use of portmanteau forms (i.e. finite verbs and attributive adjectives) which encode tense, aspect, person, and number on a single word. For verbs specifically, using periphrastic forms holds the additional advantage that the functions are separated over the distinct formal elements: whereas the finite element realises temporal and modal grounding, the non-finite element specifies the process type (Matras 2009: 182; Myers-Scotton & Jake 2014). This separation of functions over distinct formal elements in periphrastic forms may be an additional factor reducing the processing cost of loan verb accommodation.

Since accommodation biases involve the use of periphrastic strategies, we propose that biases may be linked to the processing limitations that come with loan words. That is, they exert a facilitator function to compensate for or lower the added processing cost which is imposed on loan words by increased dual-language activation. As such, loan words can be used in environments which are cognitively easier to process and produce, or in usage categories which are less hostile towards loan words, such as non-finite verb categories. This argumentation fits the characterisation of avoided structures (i.e. finite verbs, attributive adjectives, and the presence of inflection in verbs and adjectives) as storing more condensed grammatical information than their preferred counterparts (i.e. non-finite verbs, predicative adjectives, and the absence of inflection in verbs and adjectives). This explains why, in example (116), French-origin *destroien* ('to destroy') in Late Middle English is used as a bare infinitive, whereas English-origin *seien* ('to say') is used as an inflected finite form.

(116)  *as the prophete David* **seith** *, " God <u>shal</u>* **destroie** *the fruyt of the erthe "*

'As prophet David says: "God shall destroy the fruit of the earth."' (*The Parson's Tale*, c1390, PPCME2)

The argumentation that accommodation biases may aid in lowering the processing cost of loan words has been supported by three linguistic phenomena: (i) the

acquisition of non-finite verbs and predicative adjectives in neurotypical children taking place before the acquisition of finite verbs and attributive adjectives, (ii) lower-frequency items in both first and second languages being less entrenched and being more susceptible to accommodation biases than higher-frequency items, and (iii) synchronic code-switching research suggesting the same non-finite and predicative biases as found in this thesis. Although the argumentation would benefit from additional testing, it corresponds with the persistence of biases over time (cf. Section 4.4.), since they occur as long as loan words are mainly used by bilingual speakers. For them, loan words resemble cognates and can, therefore, be expected to boost dual-language activation. That is the case for, for instance, Middle English, where biases are first attested in Early Middle English and persist into Late Middle English, when contact between Anglo French and Middle English was ongoing. We did this by tracking a set of early French loans accommodating to Middle English. The biases are seen to largely persist throughout the Middle English period, and only weaken after the influx of loan words had subsided. Thus, as long as contact persisted, French lexical material continued to receive differential treatment from language users. This explanation also complies with biases weakening with higher lemma frequencies, as could be observed from the logistic regression models in Chapters 3. and 4.: the more frequent the item, the more entrenched and the easier to automatically process it is, which means that the processing cost is lower. In items with a lower processing cost, language users may have less of a need to resort to certain structures; conversely, in items with a higher processing cost, language users may have more of a need to resort to certain structures. This ties in with accommodation biases in Middle English being present as of the first attestations in contact (cf. Section 4.4.). Moreover, earlier French loans are even more common in uninflected forms than later French loans, meaning that the markedness bias in French-origin adjectives is significantly stronger in earlier than in later loans. The implication is that the processing cost coming with earlier loans may be even higher than with later loans due to lower levels of conventionalisation (as defined by Schmid 2015) in the first attestations. Therefore, with earlier loans there may be an even greater need to offset the processing cost and thus to resort to accommodation biases than with later loans.

## 5.4. Accommodation biases in relation to borrowing and code-switching

In the previous sections we have argued that loan word accommodation biases exert a facilitator function in lowering the processing cost of loan words. This deepens our understanding of the relationship between two other practices of inserted other-language material traditionally distinguished in the language contact literature,

namely (lexical) borrowing and code-switching. Both phenomena are well-documented by theoretical linguists, sociolinguists, psycholinguists, and anthropologists, yet researchers do not unequivocally agree on definitions of borrowing and code-switching as well as on their mutual relation (cf. Section 2.2.). According to Poplack & Dion (2012: 311), "distinguishing code-switching and borrowing is […] perhaps the thorniest issue in the field of contact linguistics today". For instance, are the two phenomena identical? And does one originate from the other? Those questions especially apply to single-word insertions, where the boundaries between borrowing and code-switching tend to be obscure. How we look at both phenomena can have theoretical as well as practical implications.

A first group of linguists in this controversial debate argues that borrowing and code-switching are basically different words for the same phenomenon (e.g. Coetsem 2000; Myers-Scotton 2002; Thomason 2003). According to a second group of linguists, however, borrowing and code-switching are not purely the same phenomena (e.g. Romaine 1989; Torres Cacoullos & Aaron 2003; Schendl 2004; Onysko 2007; Poplack & Dion 2012). In fact, Poplack and followers only consider borrowings (see e.g. Poplack, Sankoff & Miller 1988; Poplack, Wheeler & Westwood 1989; Poplack & Meechan 1999; Poplack 2017: 201; Poplack et al. 2020). The reason for those disagreements is that the question has been answered from different frameworks, by using different definitions, assumptions, communities of speakers, and data.

In this thesis, we follow the "researchers [who] wish to leave open the possibility that borrowings may be used by speakers in a different way from switches" (Deuchar 2005: 256). We, therefore, treat them as distinct phenomena, and provide definitions and information for both practices separately. Since borrowing has already been discussed in the state of the art, descriptions of borrowing (Section 5.4.2.) here will remain rather concise.

### 5.4.1. Code-switching

*Code-switching* (also called switching) has been defined in various ways in the language contact literature (for an overview, see e.g. Clyne 2003; Deuchar 2020). Thomason (2001: 132), for example, has defined code-switching as "the use of material from two (or more) languages by a single speaker in the same conversation". To put it more simply, language users navigate between two or more languages. It is worth mentioning that what most researchers call code-switching (e.g. Deuchar 2020) has been referred to as *code-mixing*[63] in the typological research by Muysken (2000), and as *code-alternation* by Johanson (2002: 287). Some researchers define the practice

---

[63] However, the term 'code-mixing' is by some linguists used to denote a specific type of code-switching (Zenner 2017: 237).

of code-switching as the routine alternation between languages within a continuous stretch of discourse, typically even mid-sentence (e.g. Gumperz 1982; Auer 1998; Muysken 2000). This can be illustrated by means of example (117) below, where the speaker switches fluently between Spanish and English. English-origin material is highlighted in order to contrast it to the Spanish strings.

(117)   *y este,* **he was a teacher, you know***, fue maestro y todo*

> 'And, um, he was a teacher, you know, he was a teacher and everything.'
> (Torres Cacoullos & Aaron 2003: 217)

Backus & Dorleijn (2009: 76–77) have contrasted such examples of *alternational code-switching* to *insertional code-switching*, where other-language items are inserted in a matrix language (cf. infra). The number of words in a switch is a controversial topic (Deuchar 2022): it can constitute single-word (also called word-level or isolated) insertions or multi-word insertions in a matrix language. According to Onysko (2007) and Poplack (2018), code-switches are typically multi-word stretches, which is what distinguishes them from borrowings, which are single-word lexical items. Other authors argue that single-word insertions can be code-switches, too, and that the number of words is not a valid factor for distinguishing between borrowing and code-switching (e.g. Deuchar 2005: 256). In this thesis both single- and multi-word insertions can be considered code-switches; however, the focus of this chapter is only on single-word insertions, since it is here that the boundaries between code-switching and borrowing become blurred. More concretely, the focus is on the *intrasentential*[64] (i.e. within the same sentence) use of insertions (Poplack 1980) from an embedded language into a matrix language, with a focus on switches between only two languages. In her Matrix Language Frame (MLF) model, which is a seminal model of insertional code-switching, Myers-Scotton (1993a) has defined the matrix and embedded languages as follows:

> the language with the most critical grammatical contributions to the bilingual clause is called the Matrix Language (ML), and the other participating language[65], which largely supplies some content elements in the clause, is called the Embedded Language (EL) (Myers-Scotton & Jake 2014: 3).

Typically, it is the language of the finite verb which is the matrix language. The MLF model implies an asymmetric situation between the two languages or varieties. Also, according to this model code-switching cannot occur at levels lower than the sentence

---

[64] Intrasentential switching is contrasted to *intersentential switching*, where switching happens between sentences (Poplack 1980). Poplack (1980) also distinguishes *tag-switching*, where a tag from one language is used in another language.
[65] "[P]articipating language" in this citation is used in the singular, but we acknowledge that code-switching can involve more than two languages.

(Myers-Scotton 1993a). An example of an intrasentential code-switch from English (matrix language) to Hebrew (embedded language) is shown in (118).

(118)  *It's a very complex* **yeri'á** *that he* **porés**

‘He is providing a very complex account (lit. It is a very complex sheet that he spreads.).’ (Matras 2009: 131)

Hebrew *yeri'á* means ‘sheet’ and *porés* is the present third person singular of ‘to spread’, hence ‘spreads’. The Hebrew idiom *p.r.s. yeri'á* used in (118) literally means ‘to spread a sheet’ and can be loosely translated in English as ‘to provide a complex account’. However, depending on the terminology used one could also argue that these are two examples of intraclausal switching, where a language user switches within one single clause. *Intraclausal switching* (cf. *insertion* in Muysken 2000: 60) can be opposed to *interclausal switching* (cf. *alternation* in Muysken 2000: 96), where a language user switches between clauses (Deuchar 2020). This classification refines the original formulation of the MLF model (Myers-Scotton 1993a), which looks at code-switching at the sentential or discourse level only, since more recent data have shown that code-switches can occur at the clausal level as well.

Code-switching research has proposed that the inserted material from both (or all) languages is internally complex and non-conventionalised in the matrix language (e.g. Turan et al. 2020: 1044). As such, both languages as well as their corresponding linguistic systems are activated in the language user's mind (cf. Torres Cacoullos & Aaron 2003). It is thus not surprising that code-switches are typically uttered by highly competent multilingual speakers, whose minds store the linguistic systems of several languages (Backus & Dorleijn 2009: 76–77)[66]. According to Torres Cacoullos & Aaron (2003), code-switches tend to maintain their source-language structures even in their matrix language, as opposed to borrowings, which are fully integrated into their recipient language. Many other researchers use morphosyntactic integration as a criterium to distinguish between code-switching and borrowing (e.g. Poplack 1980; Torres Cacoullos & Aaron 2003: 294; Gardani 2008: 20); however, the use of integration as a criterium has also been heavily criticised (e.g. Poplack 2000: 221–223; Deuchar 2005: 256). Although we acknowledge that code-switched material can be morphosyntactically integrated in its matrix language, integration has consistently and prototypically been linked with borrowing, and not with code-switching. Therefore, we use morphosyntactic integration as a (rather artificial) factor to distinguish between borrowing and code-switching, and we characterise borrowing by integration and code-switching by non-integration.

---

[66] However, in this thesis cases of less proficient multilingual speakers switching between languages are defined as code-switching as well.

## 5.4.2. Borrowing

The phenomenon of (lexical) borrowing typically constitutes single-word lexical items which have been copied from a source language into a recipient language at some point in the past. As a result, the lexical item starts belonging to its recipient language while continuing to exist in its source language. Internally simple material is typically integrated morphosyntactically (Torres Cacoullos & Aaron 2003: 294) to the extent where it becomes highly conventionalised in its recipient language. In fact, conventionalisation is even a requirement for a word to become a loan word (Rohde, Stefanowitsch & Kemmer 2000: 5). The material also adopts recipient-language morphosyntactic (Fischer 2007: 27; Backus & Dorleijn 2009: 77) and phonological behaviour (Fischer 2007: 27)[67]. Since borrowed material can generally be treated just like recipient-language material (cf. Wohlgemuth's 2009 direct insertion), and knowledge of the source language is not required, it follows that borrowings can be used by both bilingual and monolingual speakers (cf. Fischer 2007: 27). An example of a borrowing from English into Dutch is displayed in (119).

(119)   *Als iemand nu gewoon even 'Nee' kan* **replyen**.

    'If only someone could now quickly reply 'No'.' (Twitter, 10 November 2022)

Although loan verb *replyen* in (119) is borrowed from English *to reply*, it has accommodated to the Dutch language and it is used like any regular native Dutch infinitive, namely with an *-en*-marker attached to the stem. However, this thesis has shown that, although borrowings can generally be used like native words, they are also subject to loan word accommodation biases, which means that they are not treated identically to their recipient-language equivalents. In example (119), for instance, loan verb *replyen* is used in the infinitival form, which is a non-finite form and the preferred form for loan verbs.

## 5.4.3. Relation between borrowing and code-switching

Distinguishing between code-switching and (lexical) borrowing is still relatively controversial, and one of the points of debate lies in where to draw the line between both practices. This is a delicate point, since much is dependent on the definitions of the notions and the frameworks used. For example, Onysko (2007: 36) states that "[i]n multilingual environments, the definition of borrowing overlaps with the phenomenon of codeswitching", but this does not apply to all linguistic frameworks. Many linguists are confronted with the fuzzy boundaries (e.g. Gardani 2008: 20; Thomason 2001: 133) between borrowing and code-switching when analysing their data, encountering many "in-between" (Stammers & Deuchar 2012: 630) examples which are hard to categorise. One out of many examples which can be found in this

---

[67] However, phonological behaviour falls beyond the scope of this thesis (cf. Section 2.3.).

grey area is the phraseological borrowing or sequence (Matras 2009; Zenner & Geeraerts 2015), which is a semi-fixed multi-word foreign expression, such as *oh my god* (Zenner 2017: 238), and for which it is hard to determine whether it is a borrowing or a code-switch.

Some researchers claim that phonological integration can be used to investigate whether a switch is a borrowing or a code-switch (e.g. Poplack 1980; Besset 2017; Deuchar 2020; Gosselin 2021, 2022; Gosselin & Manning 2022). However, according to Torres Cacoullos & Aaron (2003: 290) and Poplack et al. (2020: 152) phonological integration is not a valid factor, for instance because both contact languages hare similar phonetics (Sankoff, Poplack & Vanniarajan 1990: 73)[68]. Other linguists, such as Stammers & Deuchar (2012), prefer to follow Muysken (2000: 71) by looking at *listedness*, or "the degree to which a particular element or structure is part of a memorized list which has gained acceptance within a particular speech community". This presupposes a separate mental lexicon for both the source and recipient language. If the word is listed in the recipient-language dictionary, it is considered a borrowing; if it is not listed, it is considered a code-switch (Stammers & Deuchar 2012: 642). This means that borrowings belong to two lexicons (i.e. source- and recipient-language lexicons), while code-switches belong to only one lexicon (i.e. source-language lexicon). Another factor regularly used to distinguish between borrowing and code-switching is the speaker's proficiency level: Thomason (2001: 133), for example, claims that when monolingual speakers use a foreign word in their native language, it must be a lexical borrowing, as the speaker does not know the words in that foreign language. However, this explanation cannot distinguish between borrowing and code-switching in bilingual speakers, who know both the native and the foreign language. For researchers who propose that borrowing and code-switching are different processes, it is still not always clear where to draw the line between the two phenomena: when does an inserted item stop being a code-switch, and when does it start being a borrowing? This may be testable by means of neuroscientific techniques, as proposed by Stammers & Deuchar (2012). In some cases, other-language material does not clearly belong to either borrowing or code-switching, or shows features of both. This is illustrated in example (120), reproduced from the Ottawa-Hull corpus, where English-origin material (*coper*) is used in Canadian French.

(120) *Je serais pas capable de **coper** ([kɔ'pe]) avec.*

'I couldn't cope with it.' (Poplack, Sankoff & Miller 1988: 52)

---

[68] Although Poplack was initially the linguist who introduced phonological integration as a distinguishing factor between borrowing and code-switching, she later rejected it.

In the example, English *to cope* has been morphosyntactically integrated into the Canadian French variety since it carries a French infinitival *-er*-marker, hence *coper*. Such instances provide the case for 'nonce borrowings'[69] (also called lone other-language item), a term introduced by Poplack and colleagues for words which occur only once in a body of texts (see e.g. Poplack, Sankoff & Miller 1988; Poplack, Wheeler & Westwood 1989; Poplack & Meechan 1999; Poplack 2017: 201; Poplack et al. 2020). Despite their morphological and syntactic integration (Sankoff, Poplack & Vanniarajan 1990: 94), their low usage frequency in the recipient language and the fact that they are not conventionalised distinguishes nonce borrowings from established borrowings: nonce borrowings "are not necessarily recurrent, widespread or recognized by host language monolinguals" (Sankoff, Poplack & Vanniarajan 1990: 71). Although nonce borrowings resemble code-switches in this aspect, they are also different from code-switches, for instance in that they are not internally complex.

Poplack, Sankoff & Miller (1988) regard nonce borrowings, such as English-origin *coper* in French, as borrowings since they are treated as recipient-language material. This is, among others, because the nonce borrowings receive recipient-language inflections (cf. Adalar & Tagliamonte 1998; Torres Cacoullos & Aaron 2003). They are, therefore, treated as recipient-language items. However, according to other researchers (e.g. Myers-Scotton 1993b; Jake & Myers-Scotton 2002), nonce borrowings can also be regarded as single-word code-switches, the embedded language providing a word, and the matrix language providing the grammatical template and inflectional markers for the word to be inserted into. Single-word code-switches can only begin to be considered borrowings when they start to be used more frequently (Myers-Scotton 1993b). However, Zenner & Geeraerts (2015: 250), for instance, have called into question what 'occurring only once' actually means: what is the time frame, and how many speakers are involved? Additionally, we suggest that it depends on the size of the corpus — that is, in cases where data are extracted from a corpus. The split of theoretical explanations of (and criticism on) nonce borrowings reveals a deep-rooted layer of the controversy on borrowing and code-switching, namely how they relate. One stance is that borrowings originate from single-word code-switches (e.g. Myers-Scotton 1993b), which happens through conventionalisation (cf. Backus 2005: 318 and other usage-based approaches). Since the only way for borrowings to develop is through code-switches, the two phenomena are said to be naturally linked. An opposing stance is that borrowings develop from nonce borrowings increasing in use, and that there is no connection with the phenomenon of code-switching whatsoever. From this stance it follows that borrowing and code-switching are two unrelated phenomena, as claimed by Poplack

---

[69] However, other researchers would call such instances code-switches.

and colleagues (e.g. Poplack, Sankoff & Miller 1988; Poplack, Wheeler & Westwood 1989; Poplack 2017; Poplack et al. 2020). However, if borrowings develop from nonce borrowings, the questions arise how large the overall dataset is, and what the cut-off point for frequency is.

The existence of loan word accommodation biases challenges some elements of the persistent debate on the dichotomy between borrowing and code-switching. Biases show that — despite common consensus (e.g. Wohlgemuth 2009) — even highly conventionalised and morphologically integrated borrowed material is not entirely treated like recipient-language material. Loan word accommodation biases, therefore, directly challenge Poplack et al.'s argumentation that nonce borrowings are morphosyntactically indistinguishable from recipient-language lexical items. Also, in structural terms, the presence of loan word accommodation biases aligns with some of the probabilistic morphosyntactic constraints on code-switching, such as inserted other-language verbs being more common in non-finite than in finite verb forms (Myers-Scotton & Jake 2014; also see Keller 2020: 66). The presence of loan word accommodation biases at first sight seems more compatible with Myers-Scotton's (1993) stance that borrowing and code-switching are related phenomena and that conventionalised borrowings originate from code-switches (cf. Backus 2005: 318 and other usage-based approaches). However, it also shows that even conventionalisation cannot entirely disconnect inserted items from their source language in language users' minds. For that reason, even highly integrated borrowings may not be that different from single-word code-switches, which means that the boundaries between code-switching and borrowing are fuzzy.

Assuming that borrowings can develop from single-word code-switches, as proposed by Myers-Scotton (1993b), we propose that loan word accommodation biases may characterise the gradual transition from code-switching to borrowing, possessing features of both practices. On the one hand, inserted material in accommodation biases is still recognised as source-language material by bilinguals, and its behaviour is probabilistically divergent compared to recipient-language material. In this respect, the use of the inserted material resembles a minor code-switch. On the other hand, the inserted material can have a high usage frequency, be conventionalised and be morphosyntactically integrated in its recipient language, in which it resembles typical borrowings. That biases tend to persist even in seemingly well-integrated inserted items suggests a long and gradual phase from code-switches to completely integrated borrowings, where — over time — inserted items become increasingly conventionalised and integrated in their recipient-language grammar. Inserted items only gradually develop recipient-language-like distributional properties, and they only lose all association to their source language by the time they have become fully integrated borrowings. The question remains, however, when exactly inserted items

lose their connection to their source language, and at what point they are fully integrated and show recipient-language distributional tendencies. Based on this argumentation, we suggest that the phenomena of borrowing and code-switching are naturally linked, ascribing to the usage-based claim that borrowing and code-switching are ends of the same continuum or cline, and are thus strongly related phenomena rather than two separate phenomena. This argument has previously been developed by, among others, Myers-Scotton (1993b: 176), Treffers-Daller (2005), Matras (2009), and Backus (2015). For Matras (2009: 110), the reason for viewing code-switching and borrowing as related phenomena is that "we can assume that contact-induced language change begins with the use of items from another language in conversation by people who are either bilingual[70] or have at least some exposure to another language", which is the case in both language contact phenomena. Although Stammers & Deuchar (2012) agree that there might be a continuum of integration, according to them it is unclear whether the continuum goes from code-switching to borrowing at all. Our stance lets go of the strict dichotomy between borrowing and code-switching, and accepts that boundaries between both phenomena can be fuzzy (cf. among others Thomason 2001: 133; Gardani 2008: 20; Matras 2009; Boas & Pierce 2011; Stammers & Deuchar 2012: 630; Backus 2014: 103; Zenner & Geeraerts 2015: 248; Zenner 2017: 238; Zenner, Backus & Winter-Froemel 2019: 4).

## 5.5. Possible alternative causes of accommodation biases

That we have attributed the existence and persistence of loan word accommodation biases to the elevated processing cost coming with loan words does not exclude that biases may additionally be accounted for by other factors. For instance, although the word order in complex noun phrases (e.g. *those five green cars*) can be realised in many different ways, information theory research on different languages[71] has revealed a consistent preference for a limited set of word order possibilities (for evidence, see references in Culbertson, Schouwstra & Kirby 2020). More concretely, adjectives — compared to determiners and numerals — tend to be placed as closely to nouns as possible, hence *those five green* **cars**. Culbertson, Schouwstra & Kirby (2020) have observed through experimental research that such preferred noun phrase orders also apply to improvised sequences of gestures produced by speakers of English. In a next step, they have used corpus data for several languages across language families to measure the "strength of association" (Culbertson, Schouwstra & Kirby 2020: 699) between elements in the noun phrase (i.e. adjective, demonstrative, noun, and numeral). Strength of association can be tested by means of "pointwise mutual

---

[70] However, in this thesis we recognise that other-language items can be inserted by people who are not necessarily bilingual.

[71] That is, languages with a clear category of adjectives and nouns.

information" (abbreviated PMI), which reveals "whether a given pair of elements cooccur more than would be expected from their base frequency rates" (Culbertson, Schouwstra & Kirby 2020: 703). More concretely,

> [i]f wine cooccurs with the property red more often than it would in a world in which objects and properties combined freely, then this pair will have high PMI. PMI for a pair of elements will be zero when the two elements are completely independent of one another, and negative when they cooccur less than would be expected by their base rates. (Culbertson, Schouwstra & Kirby 2020: 703)

In their research, the PMI value between adjectives and nouns was highest, which means that — conceptually speaking — adjectives (properties) are most closely related to nouns (objects). This reflects why adjectives tend to be placed closer to nouns than, for example, demonstratives and numerals (cf. *those five* _green_ **cars**). Following this study by Culbertson, Schouwstra & Kirby (2020: 703), it may be valuable to calculate PMI values for our own data as well, namely in order to assess the strength of association of attributive adjectives versus predicative adjectives with the heads which they modify. We hypothesise that attributive adjectives and their heads — in terms of collocation — will collocate more strongly than predicative adjectives and their heads, since attributive adjectives are located closer to their heads than predicative adjectives, and they are more often inflected (and thus more dependent on the morphosyntactic features of their heads). If the hypothesis is borne out, it is plausible that language users feel less reluctance towards integrating loan adjectives in predicative syntactic position than in adjective syntactic position, because predicative position imposes lower degrees of association with the noun. This hypothesis may, therefore, account for the predicative bias, and for adjectives sometimes entering into their recipient language into larger phrasal units (cf. Section 4.3.2.3.). However, this hypothesis remains highly speculative and should first be tested.

Apart from the strength of association between nouns and adjectives, which may only account for the predicative bias, other causes may be at play as well. For instance, the emergence of biases may be triggered by a set of language-internal causes, and those may depend on the language pair. Another option is the difference in position between preferred and avoided contexts: whereas attributive adjectives and finite verb forms tend to be used clause-medially, predicative adjectives and non-finite verb forms tend to be used clause-finally (cf. Section 4.1.). Flamenbaum (2014: 351) has found that code-switches between English and Twi, a language spoken in Ghana, tend to occur in turn-initial or clause-final position. Additionally, Amuzu (2016: 138) has observed that English adverbs in Ewe-English code-switching occur mainly in clause-initial and -final position. In other words, the code-switching literature suggests

that switching in clause-medial position (e.g. finite verbs and attributive adjectives) is more costly than in clause-final position (e.g. non-finite verbs and predicative adjectives). This may be another reason why accommodation biases are resorted to. However, evidence from code-switching research remains tentative overall, and more research on this topic remains to be done.

The answer to the existence of accommodation biases may also lie in language production, and the fact that language users may feel hesitant to operationalise loan words in certain constructions due to source-language incompatibility. In Dutch (examples taken from Los 2016: 270), for example, some verbal collocations are separable (e.g. *pianospelen* lit. 'to piano play', meaning 'to play the piano'), while others are inseparable (e.g. *bekvechten* lit. 'to beak fight', meaning 'to squabble'). Separable verbs tend to separate their core verb from their participle, whereas both elements in inseparable verbs appear in one word. This is illustrated in examples (121) and (122) below, where both *pianospelen* and *bekvechten* are used in the third person present singular, but *pianospelen* is separable and *bekvechten* inseparable.

(121)   *Hij* **speelt piano**.

      'He plays the piano.'

(122)   *Hij* **bekvecht**.

      'He squabbles.'

Since the division between separable and inseparable forms in Dutch verbal collocations does not exist in English, language users may have doubts concerning the separability status of a loan verb such as *hanggliden*. They may doubt whether to conjugate the verb as *Hij glidet hang* (separable) or *Hij hangglidet* (inseparable), and be biased to use the infinitival form instead, for instance in the progressive construction *Hij is aan het hanggliden*[72]. Thus, we suspect that this incompatibility between the Dutch and English language may prevent language users from fully integrating a loan verbal collocation, such as *hanggliden*, in all usage categories in Dutch. Like Dutch, other languages may contain their own unique language-dependent features which complicate loan verb integration, and which cause language pairs to be less or more compatible (for compatibility between languages, see Poplack 1980; Sankoff 1998). This could be tested by following two steps: in a first step, one could make an in-depth comparison of the language structures of different language pairs. In a second step,

---

[72] Note that the grammar rules may create confusion in the past participle as well, since Dutch past participles are usually formed by means of circumfix *ge⟩...⟨d/t*, as in *ge-wandel-d* 'walked'. However, in some of the verbal collocations *-ge-* appears in the middle of the construction (for details, see Los 2016: 271), such as in the verb *parachutespringen* 'to sky dive', which takes *parachute-ge-sprongen* as its past participle instead of *ge-parachutesprongen* (cf. Boon & Hendrickx 2015).

one could experimentally measure the processing cost in the production of items from two different languages, and one could assess whether that cost is higher in language pairs characterised by higher levels of incompatibility. If that is the case, it could be claimed that loan word accommodation biases also arise through reluctance in language production which can be due to incompatibly between language pairs. Balam, Parafita Couto & Stadthagen-González (2020: 964) have also highlighted the need to include context-specific and language-external factors when studying bilingual data.

## 5.6. Discussion

In this theoretical chapter we have investigated which factors may lead to loan word accommodation biases in contact settings between similar languages. Based on Chapters 3. and 4., Section 5.1. excluded source-language features and tendencies as well as the addition of recipient-language inflectional morphology as possible causes of the existence of persistent accommodation biases in loans. The question, therefore, remained how this phenomenon can be accounted for by the processing limitations specifically associated with loan words, which do not exist in recipient-language words.

Although dual-language activation (cf. Groot & Starreveld 2015) is present in all bilingual phenomena, it is activated even more due to the presence of cognates (Li & Gollan 2018), which complicate the monitoring of language selection. We have argued that loan words for bilinguals are essentially cognates, and that loans may thus impact language selection difficulty. We have proposed that this effect on the selection of languages results in an added processing cost when integrating loan words in recipient-language syntax. Higher processing limitations for loan words — as seen for other linguistic practices, such as code-switching (Myers-Scotton & Jake 2014) — may be accounted for by accommodation biases, which function as facilitative strategies for integrating loan words in their recipient language. In other words, the added processing cost of loan words can be tempered by means of accommodation biases, where language users operationalise loans in structures which are less dense in grammatical information, such as non-finite and predicative forms (as opposed to finite and attributive forms). This explanation of loan word accommodation biases is consistent with their diachronic persistence (cf. Section 4.4.), and with them being less prevalent with higher-frequency items (cf. Chapters 3. and 4.).

The empirical evidence of loan word accommodation biases as well as the theoretical explanation developed in this chapter may deepen our understanding of a controversial point of debate in the language contact literature, namely the relationship between single-word code-switches and lexical borrowings. Whereas

some researchers (e.g. Myers-Scotton 1993b) claim that borrowings originate from code-switches, others claim that the two practices are fundamentally unrelated, and that nonce borrowings are essentially borrowings (e.g. Poplack & Meechan 1999; Poplack & Dion 2012). However, our findings call into question Poplack and colleagues' treatment of nonce borrowings as no different from entrenched borrowings. Also, the diachronic data suggest that accommodation biases are the result of a long and gradual process of conventionalisation: even integrated borrowings keep being associated with their source language long past the point where conventionalisation has taken place, which means that code-switching and borrowing are rather similar phenomena. Our subsequent argument was that persistent accommodation biases may be characteristic of the unidirectional transition from code-switching (initial stage) to borrowing (final stage). In the transition phase, inserted items become conventionalised as they are used increasingly frequently, and they gradually reach a stage of morphosyntactic integration in their recipient-language grammar. During this entire phase, inserted items keep being associated to their source language. Therefore, this finding supports a rather recent tendency in the usage-based literature, which considers borrowing and code-switching as being part of the same cline (e.g. Myers-Scotton 1993b; Treffers-Daller 2005; Matras 2009; Backus 2015). That we have referred to processing limitations in loan words as the main reason for accommodation biases to arise does not exclude that biases may be pushed by other (so far unexplored) factors as well. As explained in Section 5.5., for example, the predicative bias in loan adjectives may relate to predicative adjectives being less strongly associated with their heads than attributive adjectives. However, this should be tested (Section 5.5.). Language-dependent factors of all kinds, and incompatibility between languages, too, may complicate the proper integration of loan words when producing language. In any case, this study would gain from additional research identifying possible causes of accommodation biases.

A caveat of this chapter is that it has provided a processing explanation as to why loan words come with accommodation biases, yet the explanation is not based on any processing experiments. That is, it would be valuable to apply neurolinguistic methods to verify whether the processing cost of loan words is indeed higher than of native words. This could be tested experimentally, for instance by means of pupillometry (as exemplified in Shechter & Share 2021), since pupil dilation can be used to measure cognitive effort (see e.g. Beatty 1982). If the pupillary responses reveal increased levels of cognitive effort for loan words compared to native words, it may be concluded that loan words indeed come with an increased processing cost (based on Shechter & Share 2021), as we have claimed in this chapter. One should additionally assess the word frequencies of loan words (both in their source and recipient language) and native words using the SUBTLEX databases for Dutch and English,

which contain word frequencies based on film and television subtitles (e.g. Keuleers, Brysbaert & New 2010 for SUBTLEX-NL; Brysbaert, New & Keuleers 2012 for SUBTLEX-US). This way, frequency can be controlled for in the experiment, for example by only comparing loan and native words with roughly the same frequencies. Another potential caveat is that this chapter has referred extensively to the code-switching and processing literature, which mainly focuses on the spoken medium, while some of the material (e.g. PPCME2 and Helsinki Corpus) used in this thesis belongs to the written medium. Although code-switching can be a written phenomenon, written language differs from spoken, naturalistic conversation in that it is constructed with less of a time constraint. However, since the Dutch data are based on a spoken corpus and paint a similar picture to the Middle English data, we do not deem this problematic. Next, as in the previous two chapters, the reasoning developed in this chapter should not be overgeneralised as applying to all languages in the world. Typological research points to the tendency of linguistic research to focus mainly on the European languages belonging to the Indo-European language group (e.g. Velupillai 2012: 51). For example, code-switching research deals with an overrepresentation of studies on Spanish and English (e.g. Bellamy & Parafita Couto 2022: 11), two Indo-European languages. This results in a lack of data or documentation of other languages (e.g. Velupillai 2012: 51) and — by extension — communities. For instance, communities where code-switching is the norm as an unmarked speech register are all too often overlooked (Parafita Couto, Greidanus Romaneli & Bellamy in press: 4). An example is the language situation in Porto-Novo, the capital of Benin, where speakers switch all the time between French, English, and the Gbe languages (Aboh 2022). In fact, there is little research on such mixed language patterns, as well as the cognitive consequences of using multiple languages at all times (Thomason 2001: 218). An example of such a mixed language which emerged through "conventionalized codeswitching practices" (O'Shannessy 2015: 289) is Light Warlpiri, a language spoken by the Warlpiri community in northern Australia, which "systematically combines elements of Warlpiri (Pama-Nyungan) and varieties of English and/or Kriol (an English-lexified creole)" (O'Shannessy 2015: 289). What complicates research on code-switching even further is that code-switching can change over time in one community, as shown for Turkish-Dutch bilinguals in the Netherlands (Backus & Demirçay 2021; also see Demirçay 2017). Whereas code-switching was initially limited to the occasional insertion of Dutch words into Turkish, it has evolved into the integration of entire alternations (or switches between clauses), leading to a bilingual speech style. It can, therefore, not be assumed that the processing cost of loan words is just as high in (all stages of) all language contact situations around the world.

The present chapter concludes Part II, in which we have provided substantial empirical evidence for the existence of persistent loan word accommodation biases. Now that we have also provided more insight into the causes of accommodation biases, a logical follow-up question concerns the possible consequences of biases, which will be the topic of Part III. It consists of two hands-on chapters and discusses the long-term impact which loan word accommodation biases may have had on the history of the English language. The reason to focus solely on the French-Middle English contact setting is that this language contact setting has ended and that it has left a long legacy of diachronic data, which makes it more suitable for this type of analysis than the English-Dutch contact setting. A first possible diachronic consequence of accommodation biases in the English language is the rise of *do*-support in Early Modern English, which is the focus of Chapter 6.

# Part III

# Consequences of accommodation biases

# Chapter 6

# The rise of *do*-support in Early Modern English

## 6.1. Introduction

Part II of this thesis has dealt with the existence and persistence (cf. Chapters 3. and 4.) as well as the causes (cf. Chapter 5.) of loan word accommodation biases. In Part III, we will now discuss if — and how — loan word accommodation biases affect the grammar of the recipient language in the long run. Diessel (2007: 117), for instance, has found that "[s]mall biases in language production can lead to diachronic change". Loan words come with divergent distributional tendencies compared to native words, such as greater reliance on periphrastic structures (e.g. non-finite and predicative biases). Provided that the influx is constant and sufficiently large, the diverging distributional properties of loan words entering a recipient language may shift the distributional properties of that language in general. For example, the consistent bias of English loan adjectives in Present-day Dutch towards predicative forms may gradually extend beyond loan adjectives and start affecting all adjectives in Dutch, thus increasing the overall use of predicative adjectival position in Dutch. What is more, the biases create an environment which encourages the grammaticalization of specific markers that — in one way or another — pattern well with loan verbs. To investigate the possible long-term impact of biases on their recipient-language grammar, the focus of Part III will be exclusively on contact between French and Middle English. First, there is a long legacy of diachronic texts from different phases throughout the period of contact, with writers such as Chaucer, Gower, and Malory. Second, the contact setting comes with a clearly defined start and end point (or at least a defined period), and can, therefore, draw a complete picture of any linguistic change. In order to investigate possible consequences of accommodation biases on the development of the English language, Part III will focus on the emergence of two verbal periphrastic constructions in the history of English, namely the rise of *do*-support in Early Modern English (cf. Chapter 6.) and the rise of light verbs in Late Middle English (cf. Chapter 7.).

The use of *do*-support started rising as of Late Middle English, and it surged in 16th-century Early Modern English. Both periods are characterised by rapid change as compared to Old English and Late Modern English (Denison 2003: 68; as proven quantitatively by Nevalainen et al. 2020: 26). Late Middle English was marked by the general reduction of its inflectional apparatus "[a]s a result of the merging of unstressed vowels into a single sound" (Pyles & Algeo 1982: 153). The loss of inflectional morphology also affected verbs (e.g. Pyles & Algeo 1982: 153–154; Rissanen 2000). For instance, the *(e)n*-suffix in infinitives and present plurals, as in *walk-en*, was increasingly omitted throughout the Late Middle English period, as described in detail in Section 3.3.1.1.1. The onset of the deflection process can probably be situated before the Middle English period, and was more advanced in Northern than in Southern England. This ties in with the idea that the loss of inflectional morphology was linked to contact between English and Scandinavian (Allen 1997; McWhorter 2002; Emonds & Faarlund 2014), which took place between roughly 787 and 1042. Whereas Northern England was part of the Danelaw and was thus influenced by the Scandinavian language, Southern England has never been under Scandinavian rule, and neither has it undergone such influence (Thomason & Kaufman 1991: 280–282). Another development in Late Middle English, linked to the general deflection rates, was the drift towards an increased reliance on non-finite forms (e.g. Görlach 2003: 97; Green 2017). For example, Late Middle English saw the emergence of a set of periphrastic constructions which relegated lexical verbs to non-finite slots instead of inflecting them. Examples include the rise of modal verbs, light verbs, gerunds, infinitives, progressive constructions, and — the focus of this chapter — *do*-support.

In what follows, we will first define *do*-support as a periphrastic strategy where *do* serves as a semantically empty operator carrying inflections (Section 6.2.1.). After a brief discussion of some of the theories on the origin of *do*-support (Section 6.2.2.), Section 6.2.3. will describe the syntactic contexts and clause types where *do*-support is attested in Present-day English versus Early Modern English. More concretely, using *do*-support in Early Modern English affirmative declaratives was more frequent than not using *do*-support, whereas affirmative declaratives in Present-day English are generally used without *do*-support. Although the rise of *do*-support in affirmative declaratives has repeatedly been approached from a syntactic perspective (e.g. Kroch 1989; Kauhanen & Walkden 2018), this chapter aims to complement previous insights into the emergence of *do*-support with insights from a language contact perspective. That is to say, the increase in the use of *do*-support during the 16th century immediately followed a period of intense contact with French. This may mean that the great French influx in Late Middle English considerably boosted the usage rates of *do*-support, and that the effects became clearly visible in Early Modern English (Section 6.2.4.). This hypothesis will be tested in corpus research on

PPCEME data. However, based on the literature, an analysis of *do*-support would be incomplete without taking into account variables other than source language as well, which is why we included lemma frequency (Section 6.3.2.2.) and inflection (Section 6.3.2.3.). We equally looked at syntactic factors associated with the current use of *do*, namely clause structure (Section 6.3.2.4.), subject-verb inversion (Section 6.3.2.5.), and the presence of medial adverbials (Section 6.3.2.6.). Last, we also assessed the role of semantics (Section 6.3.2.7.). More details on the data and methodology used in this corpus study will be described in Section 6.3., and the findings will be the topic of Section 6.4. As expected, *do*-support is significantly more likely with French loan verbs than with native English verbs, which confirms the language contact hypothesis. However, the other variables play a significant role as well: *do*-support is more prevalent with lower-frequency verbs than with higher-frequency verbs, and in clauses with a non-default SVO word order than in clauses with a default SVO word order. The dataset also contains several attestations of *do*-support in contexts where it is not probabilistically expected to occur, namely with English-origin, higher-frequency verbs used in default SVO structures. This will be accounted for by the semantics of *do* (Section 6.4.7.), as a qualitative analysis will show that *do* — in such cases where it is not probabilistically expected — is often used to convey (adversative) truth emphasis. The findings, hence, reveal that multiple factors are involved in the variation between verbs with and without *do*-support. An elaborate discussion of the findings will conclude this chapter in Section 6.5.

## 6.2. *Do*-support

### 6.2.1. Definition

*Do*-support, the term which will be consistently adopted throughout this thesis, is also frequently referred to as auxiliary *do*, *do*-construction, periphrastic *do*, and *do*-periphrasis. Since the origin and historical development of *do*-support encompass a classical problem, the literature on this topic is particularly rich (cf. Warner 1993: 13), with seminal work conducted by Ellegård (1953) in the fifties. Later influential authors have been, among others, Kroch (1989)[73], Denison (1993), and Warner (1993), who have approached the topic from a generative (Kroch) and non-generative (Denison and Warner) framework.

The periphrastic construction of *do*-support consists of a form of *do* combined with a lexical verb (Warner 1993), as in *he <u>does</u> **speak***. In instances of *do*-support, a form of *do* is combined with the infinitive of the lexical verb. Following Ellegård (1953), the form of *do* is typically inflected and used finitely (*does* in the example indicating the third person singular of the present), while the lexical verb is used non-finitely and takes the form of an infinitive (*speak* in the example). Semantically speaking,

---

[73] Under the scope of this thesis, we will not expand on Kroch's (1989) Constant Rate Hypothesis.

pro-verbal *do* in this structure serves as a support auxiliary, namely an empty operator not containing a specific lexical meaning[74] (Denison 1993; De Keyser 2014; Fischer, De Smet & Wurff 2017), as it is the lexical verb which carries meaning. Therefore, *do* has also been called 'dummy *do*' (Denison 1993). Since "*do* is a highly general activity verb, roughly meaning 'perform, act'" (Fischer, De Smet & Wurff 2017: 128), and it serves as an operator carrying grammatical information regarding tense and agreement, its function resembles the function of light verbs discussed in Wohlgemuth's (2009) light verb strategy (cf. Sections 2.3.1.3. and 7.4.1.2.).

*Do*-support is a typical feature of English, and it distinguishes English from the other West-Germanic languages (McWhorter 2002: 250). However, constructions similar to *do*-support are found in many languages around the world (for a typological overview, see Jäger 2006). Examples are Danish (e.g. Houser et al. 2006), Swedish (e.g. Platzack 2008: 5), Old Icelandic (e.g. Viðarsson 2009), Scots (e.g. Gotthard 2019), and *fa*-support in the Camuno dialect spoken in northern Italy (e.g. Swinburne 2022).

## 6.2.2. Origin

*Don* ('to do') in Old English was mainly attested as a pro-form, where it resumes a word or phrase uttered earlier, as in *He does not like spiders, but I **do*** (Stein 2011: 16). The first attestations of *do*-support appeared in the written record near the end of the Middle English period, namely in a 13th-century rhyme from Southwest England (Denison 1993). However, it is possible — and even likely — that its use in spoken language had already started before that point. In the course of the Middle English period, *do* grammaticalised and fulfilled several functions typical of general activity verbs becoming grammaticalised, such as pro-verbs and emphatic markers (Fischer, De Smet & Wurff 2017: 128).

The origin of *do*-support is still a much-contested topic which has been tackled from a wide variety of frameworks. That said, according to Auwera & Genee (2002: 302), "we are much closer now than we were thirty years ago to at least being able to distinguish between plausible and less plausible scenarios". Although not universally accepted, the most commonly used hypothesis is that *do*-support developed out of the use of *do* as a causative auxiliary (e.g. Ellegård 1953)[75], two examples of which are provided in (123) and (124).

---

[74] However, it would be wrong to claim that *do* does not add any semantic value at all. For example, in Early Modern English *do* served as a marker of truthfulness (Budts 2020a, 2020b), and in Present-day English it can be used to convey emphasis (Huddleston 1976), as will be discussed in the following sections. The semantics of *do* will prove to be relevant in our case study as well (cf. Section 6.4.7.).

[75] An argument against assuming causative *do* as the origin of *do*-support is that the use of causative *do* combined with an infinitive was actually rare in Old English (Ellegård 1953).

(123)   *The Kyng..<u>ded</u> his officers **arestin**..his uncil, the duke of Gloucetir.*

‘The king had his officers arrest the duke of Gloucester.’ (*John Capgrave's Abbreviacion of Cronicles*, a1464, MEC)

(124)   *they <u>did</u> hys master to **understand** whate hys man had sayed unto them.*

‘They made his master understand what his man had said to them.’ (*The Autobiography of Thomas Mowntayne*, c1555, PPCEME)

In example (123), the officers perform the action of *arestin* ('to arrest'), but it is the subject (*The Kyng*) who makes the action of *arestin* happen by ordering his officers to do so. Therefore, *ded arestin* carries the meaning of 'had them arrest' or 'made them arrest'. Example (124), next, shows that, although subject *they* makes the understanding happen, it is *hys master* who actually performs or undergoes the action of understanding. Both examples show that the use of causative *do* is comparable to the use of causative *make*, *let*, and *have* in Present-day English (Garrett 1998: 287), and of causative *bringen* ('to bring') in Present-day German (Percillier 2022 p.c.). Syntactically speaking, it resembles the use of *do*-support, since both constructions combine a form of *do* with an infinitive. According to Ellegård (1953: 92), the development of causative *do* in English may even have been sped up by French causative *faire*, which "may have contributed indirectly to the establishment of [*do*-support]". Another explanation, suggested by Denison (1985: 52), is that causative *do* entered the language as a Latinism.

In line with Ellegård (1953), this thesis assumes that *do*-support as it is known today may have evolved from a causative marker into an auxiliary verb. An example like (125), for instance, is ambiguous in use and may have served as such a bridging context from causative to auxiliary *do*. A possible interpretation of example (125) is that *do* has been semantically bleached, and is used in a non-emphatic affirmative declarative as a semantically near-empty modal accompanying *sauen* ('to save'). However, one could also argue that *do* in the example still carries a trace of its causative meaning, namely that a soul can be saved by pardon, penance, and prayer. *Do* may later have grammaticalised into an operator to form certain clauses, as is shown in (126) below, where it is used in an interrogative clause. The use of *do* as an operator in this example from ca. 1460 corresponds to one of the uses of *do* in Present-day English.

(125)   *And so bileeue I lelly..Þat pardoun and penaunce and preyers <u>don</u> **sauen** Soules þat han sunget.*

‘And so I sincerely believe that pardon and penance and prayer do save souls who have sinned.’ (*Piers Plowman*, c1425, MEC)

(126) <u>*Do ye no*</u> **drede** *God þat is a-bove?*

'Do you not dread God who is above?' (*The Tale of Beryn*, c1460, MEC)

The other explanations for the origin of *do*-support include that *do x* is equivalent to Old French *faire x* (Ellegård 1953; Denison 1985), and that it has arisen under Celtic influence (Preussler 1956; Auwera & Genee 2002). We do not elaborate on those explanations. According to Poussa (1990), *do*-support developed as a creolisation feature of Anglo Welsh, but this explanation has generally received little support, and has been rejected by, among others, Denison (1993). This list is by no means exhaustive, with hypotheses coming from syntactic, semantic, dialectal, and socio-historical frameworks (cf. Stein 2011: 27). However, within the scope of this thesis we will focus only on causative *do* as the origin of *do*-support.

## 6.2.3. Evolution of usage

The use of *do*-support in Present-day English is mandatory in four contexts, which are illustrated in examples (127), (128), (129), and (130) below.

(127) *I <u>do</u> not* **love** *you.*

(128) <u>*Do*</u> *you (not)* **love** *me?*

(129) *So you* **ate** *all the cookies, <u>did</u> you?*

(130) *I <u>dó</u>* **love** *you!*

The above examples, which are reproduced from Budts (2020a: 1), display the acronymic 'NICE'-environments (Huddleston 1976), which are the only four syntactic contexts still associated with *do*-support in Present-day English: **n**egation (127), subject-verb **i**nversion (128), **c**oding previously mentioned material (e.g. tag questions) (129), and conveying **e**mphasis[76] (130). The use of *do*-support in Present-day English is generally restricted to those NICE-environments — which have non-default SVO word orders — and its use is never optional.

Similar to Present-day English, *do*-support in Early Modern English (1500–1720) was most common in questions[77] (i.e. negative questions and affirmative questions), followed by negative clauses (i.e. negative declaratives and negative imperatives). This can be seen from Figure 11, which has been reproduced from Budts (2020b: 4). It is based on the work by pioneering researcher Ellegård (1953: 162).

---

[76] However, *do*-support in Southwest England can be used in non-emphatic affirmative declaratives (Klemola 1998).

[77] In this thesis, we will use the term 'interrogative'.
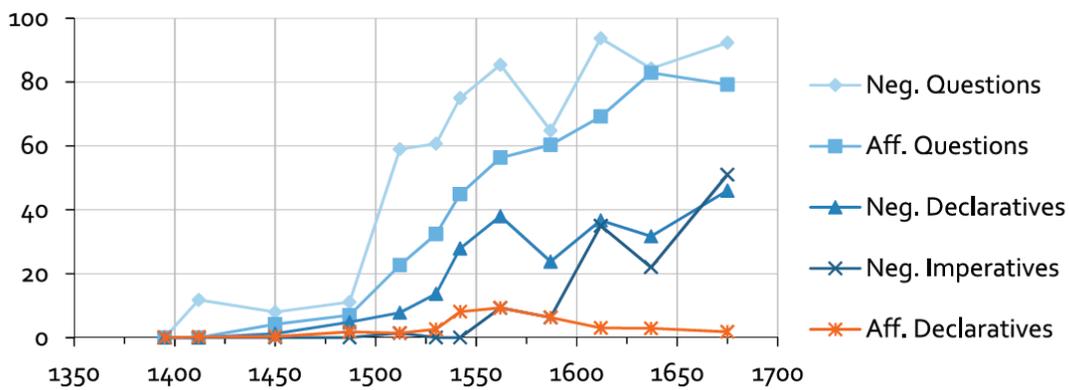
*Figure 11: Relative frequency of* do. *Adapted from Ellegård (1953: 162),*
*in the version adopted by Budts (2020b: 4).*

Different from Present-day English, the use of *do*-support in 16th-century English was optional and extended to non-emphatic affirmative declaratives, as can be seen from Figure 11. Two instances of non-emphatic affirmatives are displayed in (131) and (132).

(131)  *And this square <u>doth</u>* **contain** *the first quadrate A.B.C.D , and also a squire G.H.K*

'And this square contains the first quadrate A.B.C.D., and also a square G.H.K.'
(*The Path-way to Knowledg*, 1551, PPCEME)

(132)  *As thou <u>dyddest</u>* **send** *me into the worlde , even so have I sent them into the worlde*

'As you sent me into the world, so did I send them into the world.' (*The New Testament*, 1534, PPCEME)

Another difference with Present-day English is that *do*-support in Early Modern English was not mandatory and co-existed with non-support. Figure 11 does not illustrate that affirmative contexts were in fact much more frequent than the other syntactic contexts. We, therefore, included a plot reproduced from Budts (2020b: 4) with the normalised frequencies of *do* per thousand finite clauses (cf. Figure 12). It shows that the environment in which *do*-support was most common is the affirmative declarative, with a peak in usage around 1550, and that the usage of *do* in affirmatives quickly started decreasing after that date (Ellegård 1953).
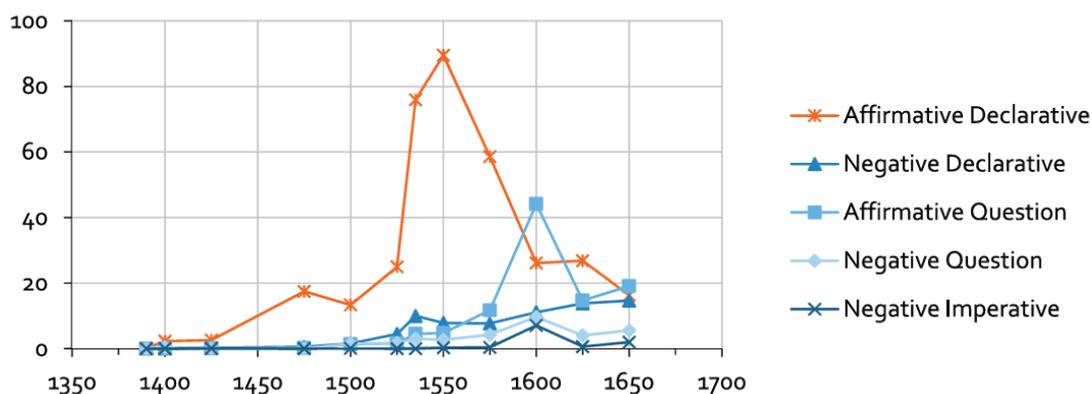
*Figure 12: Normalised frequency of* do *per thousand finite clauses. Based on counts in Ellegård (1953: 161–162), in the version adopted by Budts (2020b: 4).*

Interestingly, the non-emphatic affirmative declarative, a context where *do* ceased to be operational in Present-day English, was the environment in which *do*-support initially emerged in the late 13th century, and where it peaked in 1550 (cf. Figure 12).

The use of *do* in 16th-century affirmative declaratives has been investigated by Budts (2020a, 2020b), who has used Convolutional Neural Networks to show that *do* functioned as an epistemic modal verb. As an epistemic modal verb, it imposed a sense of truthfulness onto the proposition, as in example (133).

(133)   *And if the said T.N. his heirs, executors or administrators,* <u>*do*</u> **fail** *or make default, and do not well and truly acquit, discharge or save harmless the said T.S. G.F. [...]*

‘And if the said T.N.’s heirs, executors, or administrators fail or commit an offense, and do not well or truly acquit, discharge, or harmlessly save the said T.S. G.F. […].’ (Budts’ *Antigoon Corpus*, emphasis and translation added)

The example, which is reproduced from Budts (2020a: 166), shows how *do* is used as an empathic marker in a legal text, where it “adds a layer of truthfulness and objectivity, which is rhetorically desirable in official documents and legislation” (Budts 2020a: 166). As such, *do* in Early Modern English followed the paradigm of modals (the canonical ones being *can*, *may*, *must*, *shall*, and *will*) and auxiliaries. This can also be seen from the fact that pro-verbal *do* cannot formally coexist with other auxiliaries or (semi-)modals (Warner 1993: 152), as in *\*you* <u>*do*</u> **should** or *\*he* <u>*does*</u> **be**. Although modal auxiliaries in the English language had already grammaticalised in the 12th century (Denison 1993; Warner 1993: Ch. 8.), they kept developing until 1550 (Budts 2020a, 2020b), which is when the use of *do*-support in affirmative declaratives surged (cf. Figure 12). Since *do* was used as a modal verb marking tense and mood, Budts (2020a, 2020b) has suggested that the strong rise of *do* in Early Modern English affirmative declaratives may have been caused by analogy with the modals.

A possible cause of the rise of *do*-support has been discussed by Fischer, De Smet & Wurff (2017: 130): "the rise of *do* systematized an existing statistical trend [namely for negative and interrogative clauses in Middle English to occur with an auxiliary] that already distinguished negatives and interrogatives from declaratives". As a result, this pattern spread to all negatives and interrogatives, making the system more uniform. Additionally, it is probable that the rise of *do* interacted with word order changes in Middle English (Warner 1993; Fischer, De Smet & Wurff 2017). For instance, whereas verb-subject order had always been quite common in English in interrogatives as well as declaratives, subject-verb order took over as the dominant order over the course of the Middle English period, following the loss of verb-second. *Do* can be thought of as a solution to preserve a form of inversion in interrogatives, while at the same time keeping the subject before the lexical verb, as illustrated in examples (134) and (135). Both instances have been cited from Fischer, De Smet & Wurff (2017: 131).

(134) V        S        O

   ***Saw***   *we*      *a unicorn?*

(135) V-aux        S        V-lex        O

   <u>*Did*</u>        *we*        ***see***        *a unicorn?*

The use of *do*-support in example (135) allows for the innovative subject-verb word order (*we see*). Indeed, Fischer, De Smet & Wurff (2017: 131) have suggested that the use of *do* may have acted as a strategy to "bring interrogatives back in line with an increasingly strict SVO grammar".

Another example concerns adverbs (such as negative adverb *not*), which had previously been placed behind the verb, but which started being placed between the subject and the verb as of Middle English. This change of adverb placement could be realised by means of *do*-support. Example (136) shows the old use, whereas example (137) depicts the new use with *do*-support, and both examples have been reproduced from Fischer, De Smet & Wurff (2017: 131).

(136) S        V        Adv        O

   *We*   ***saw***   *not*        *a unicorn* (emphasis added)

(137) S        V-aux    Adv    V-lex    O

   *We*   <u>*did*</u>        *not*        ***see***        *a unicorn* (emphasis added)

Other researchers as well have linked the rise of *do*-support to the loss of verb movement (for references, see Haeberli & Ihsane 2016). From a generative angle, Haeberli & Ihsane (2016) have provided evidence of verb movement in structures

with adverbs. For instance, as of the 16th century V*not* (e.g. *he **spoke** not*) was replaced by *do*-support (e.g. *he **did not** speak*), and this trend may have contributed to the peak of *do*-support.

In her book chapter on the rise and regulation of periphrastic *do* in 16th- and 17th-century English, Nurmi (2000) has investigated the role of sociolinguistic variables connected with the developments, such as the role of gender. Interestingly, *do* in the 17th century was mainly used by women, but this finding postdates the period under investigation in this thesis and will, therefore, not be expanded on.

### 6.2.4. Hypothesis

As described in Section 6.1., Late Middle English saw the emergence of a number of periphrastic constructions in which the lexical verb could be used non-finitely. This development more or less coincided with the peak of French influx, with borrowing rates as high as 30% during the second half of the 14th century (Mugglestone 2006; Baugh & Cable 2013). French loans — and more specifically the more recent ones — typically entered the English language as low-frequency items, which generally adopt newer syntactic variants and patterns available in the language (Bybee & Hopper 2001). Consequently, it is plausible that French loan verbs were more likely to be used with innovative *do*-support. What is more, French loan words in Middle English were generally avoided in finite forms (cf. Chapter 3.), and *do*-support could act as a tool in doing so. We, therefore, hypothesise that French-origin verbs entering into English may have been substantially more frequent with *do*-support than English-origin verbs. The dramatic rise of *do*-support in Early Modern English may have been a delayed effect of the French influx in Late Middle English, since Early Modern English immediately followed the Late Middle English period, and any effect of French influx presumably had become demonstrable by then. In other words, the great amount of French influx (i) may have promoted the grammaticalisation of new periphrastic strategies, such as *do*-support in Late Middle English, and (ii) may have significantly triggered the reliance on existing periphrastic strategies, such as *do*-support in Early Modern English. This is an example of how accommodation biases in French loan words may have *indirectly* impacted English syntax.

Support for this hypothesis has been found in a variety of sources. In typological research, for example, authors such as Jäger (2004: 10) have found that *do*-support-like constructions are adopted as a cross-linguistic strategy to accommodate loan verbs:

> The use of an inflected auxiliary in a large number of verb constructions provides an avoidance strategy for morphological complexity or uncertainty in the case of borrowed stems, since the speaker only has to acquire a single paradigm, namely that of the 'do'-auxiliary.

An example is the use of *hata* ('to do') in Korean with Chinese and Japanese loans entering the language (Auwera & Genee 2002: 296), and the use of *dir* ('to do') in Moroccan Arabic with Dutch loan verbs (Boumans 2007). This tendency to facilitate loan word integration is in line with Moretti (2021, subm.: 25), who has stated that "the possibility to use a semantically neutral device like auxiliary *do* served to facilitate the adoption of new vocabulary items". It is also consistent with code-switching research, which has shown that many embedded-language verbs are integrated in their matrix language by means of *do* (Myers-Scotton & Jake 2014: 11–12). This tendency relates to the non-finite bias found in code-switching (cf. Section 5.3.3.), and it decreases the "processing cost associated with language switching" (Myers-Scotton & Jake 2014: 11). The above evidence supports our claim that *do*-support can simplify the process of loan verb integration: the loan verb does not need to be inflected, so the morphological complexity is lowered.

Corroborations from the literature on English historical linguistics are scarce. To our knowledge, none of the work on the history of the English language has yet connected the steep rise of *do*-support to the large influx of French words. Moreover, the language contact hypothesis has even been rejected, for instance by Hall (1983: 123): "we will simply dismiss the claim that French influence is to be seen as the specific cause of the development of [periphrastic *do*]". An exception is the work by Fischer & Wurff (2006: 155), who have proposed that speakers of Middle English may have used French loan verbs in *do*-support as a tool to avoid adding English inflections. However, they have not provided any data or further references to strengthen this claim, which remains to be tested. Moretti (subm.: 3) has stated that one of the functions of auxiliary *do* in Middle English "was to support the use of verbs of foreign origins and uncommon, low frequency verbs", which is in line with our hypothesis. Indeed, loan words and low-frequency items put a burden on processing and hence increase the processing cost (cf. Section 5.3.4.). Therefore, French loan verbs in Middle English may have been slotted in *do*-support significantly more often than native English verbs.

Last, first-language acquisition has also demonstrated that "light verbs such as copulae, modals, and auxiliaries are used in finite constructions earlier than lexical verbs" (Verhagen 2009: 203). This implies that it is easier to inflect *do* in *do*-support than lexical verbs, and it is line with our hypothesis that French loan verbs — which are associated with an increased processing cost — are favoured in *do*-support.

In order to gauge the effect of source language (Section 6.3.2.1.) on the variation between the presence and absence of *do*-support, we will now present a case study on Early Modern English corpus data. Although French influx may have partially triggered the peak of affirmative *do* in 16th-century English, it has developed within a fast-changing linguistic climate and may thus also be dependent on other variables.

Based on our previous research and the literature, the variables discussed in the present chapter are lemma frequency (Section 6.3.2.2.), inflection (Section 6.3.2.3.), clause structure (Section 6.3.2.4.), subject-verb inversion (Section 6.3.2.5.), medial adverbials (Section 6.3.2.6.), and semantics (Section 6.3.2.7.). The variables, as well as their corresponding substantiated hypotheses, will be introduced in Section 6.3.2.

## 6.3. Data and methodology

### 6.3.1. Data extraction

Data for this diachronic case study were extracted from the *Penn-Helsinki Parsed Corpus of Early Modern English (version 2)*, abbreviated PPCEME (Kroch, Santorini & Delfs 2004), a sub-corpus of the *Penn Parsed Corpora of Historical English* (Kroch 2020). This 1.7-million-words diachronic corpus covers prose from between 1500 and 1720, but for this case study we focused solely on the first sub-period, called "E1" (1500–1569), which is when the use of *do*-support surged dramatically. The corpus contains a total of 448 text samples.

In a first step, attestations of *do*-support were collected from the first sub-period in PPCEME, using part-of-speech tags "DOP" and "DOD" for attestations of present (e.g. *he does* **talk**) and past tense *do* (e.g. *he did* **talk**) respectively. Attestations of *do* were limited to a maximum of 25 from any single text in order to avoid possible textual idiosyncrasies dominating the overall dataset. In a second step, each attestation of *do* in our sample was randomly matched to an attestation without *do* occurring in the same text. Attestations without *do* were lexical verbs used in the present (e.g. *he* **talks**) or past tense (e.g. *he* **talked**), as identified by the part-of-speech tags "VBP" and "VBD" respectively. We selected only those *do*-less contexts where *do* was actually an option in order to respect the envelope of variation between the presence and absence of *do*. The advantage of using a text-matched sample is that it automatically controls for any text-level effects relating to genre, register, and style. This is desirable since formal and literary genres have been linked to an increased usage of *do*-support (e.g. Ellegård 1953; Stein 1990)[78].

However, some false positives unavoidably sneaked into the dataset and had to be excluded manually. Each false positive was replaced by a randomly selected attestation from the same text and of the same type (i.e. with *do* or without *do*), provided such an attestation was available. If no replacement was available, both the false positive and the attestation paired to it were removed from the dataset, thereby preserving the balanced proportion of attestations with and without *do*. Such false positives included cases which are not part of the envelope of variation between the presence and

---

[78] Despite the popular suggestion that *do*-support is associated with higher registers, authors such as Tieken-Boon van Ostade (1990) and Rissanen (1991) have proposed that *do*-support established itself in (informal) spoken language.

absence of *do*-support, such as pro-form *do*, where *do* is not combined with a lexical verb but resumes a word or phrase mentioned earlier in the sentence. In example (138), for instance, *dyd* is a past pro-form resuming *punyshe synne*: instead of repeating *as he **punyshed synne** then*, author Hugh Latimer writes *as he **dyd** then*.

(138)  *he wyl **punyshe synne** as well nowe as he <u>dyd</u> then*

'He will punish sin now as he did then.' (*Sermon on the Ploughers*, 1549, PPCEME)

We also excluded attestations of lexical *do*, where *do* is used transitively. An example is presented in (139), where *a murder* is the direct object of *did*.

(139)  *A Frenchman was sent againe into Fraunce to be delivered againe to the Frenchmen at the borders , bicaus of **a murder** he <u>did</u> at Diep*

'A Frenchman was sent again into France to be delivered to the French at the border because of a murder he committed at Dieppe.' (*The Diary of Edward VI*, 1550–1552, PPCEME)

Since causative *do* (cf. Section 6.2.2.) is a subtype of lexical *do*, it was equally excluded from data analysis. Last, we removed all attestations of verbs resisting *do*-support, such as *will*, *would* (in the sense of *wish*), *quoth*, and impersonals (e.g. *me **liketh**, me **thinketh***). After manual data cleaning, our sample contained a total of 1,592 relevant attestations, namely 796 attestations with *do* and 796 attestations without *do*. 959 attestations are of English origin (60.2%) and the remaining 633 attestations are of French origin (39.8%).

## 6.3.2. Data annotation

To test which variables affect the presence or absence of *do*-support (dependent variable), all attestations were manually annotated for the following independent variables: source language (Section 6.3.2.1.), lemma frequency (Section 6.3.2.2.), inflection (Section 6.3.2.3.), clause structure (Section 6.3.2.4.), subject-verb inversion (Section 6.3.2.5.), the presence of medial adverbials (Section 6.3.2.6.), and semantics (Section 6.3.2.7.). Annotations were finalised in inter-annotator agreement with Hendrik De Smet. Annotation categories and hypotheses for all seven variables will be discussed below.

### 6.3.2.1. Source language

As explained in Section 6.2.4., we hypothesise that the rise of *do* in Early Modern English may have been linked to the borrowing peak of French items in Late Middle English. French items typically entered into Late Middle English as low-frequency items, which are generally described as innovative (Bybee & Hopper 2001) and thus align with new syntactic developments and variants in the language. The rise of

*do*-support may, therefore, correlate with French influx and the tendency of French loan verbs to be used in innovative *do*-support, where finite forms can be avoided.

To test this hypothesis, the source language of the lexical verb (in bold) was manually annotated for either English or French origin, such as English-origin *make* in (140) and French-origin *deny* in (141).

(140)    *How longe <u>dost</u> thou* **make** *vs doute ?*

'How long will you make us doubt for?' (*The New Testament*, 1534, PPCEME)

(141)    *The bishop of Winchestir <u>did</u>* **deny** *the articles that the bishop of London and the other had made .*

'The bishop of Winchester did deny the articles that the bishop of London and the other had made.' (*The Diary of Edward VI*, 1550–1552, PPCEME)

Following the identification method of source language as used in Chapters 3. and 4., attestations of French origin included verbs of shared French and Latin origin. Note that borrowing from Latin in the Early Modern English period was so strong that it even

exceeded French in the period, and that the influx of Latin loan words peaked[79] between 1575–1675, with more than 13,000 loan words entering the lexicon during that time. (Moore 2017: 298)

The extensive borrowing from Latin in the 16th century tied in with "the Renaissance ideal of expressing an idea in as many ways as possible" (Nurmi 2017: 23), resulting in loan words such as *providence* and *prudence* (Cowie 2017: 60). Under the scope of this thesis, however, verbs of purely Latin origin were not considered in the analysis.

### 6.3.2.2. Lemma frequency

Chapter 3. has revealed that the non-finite bias for French loan verbs in Late Middle English was even stronger when the frequency of the verbs was lower. We have linked this tendency to the difference in levels of entrenchment between lower- and higher-frequency items (Section 3.3.2.1.; cf. Langacker 1987; Bybee 2010; Schmid 2016), and showed in Section 5.3.2. that lower-frequency items come with a higher processing cost than higher-frequency items (e.g. Berglund-Barraza et al. 2019; Desai, Choi & Henderson 2020). Since the use of *do*-support essentially involves the use of non-finite forms, the possibility of frequency effects cannot be ruled out in this

---

[79] However, some authors have claimed that the massive borrowing peak from Latin in the 16th century is due to the sampling distributions in the OED for the 16th century compared to the 18th century (for more explanations and references, see Cowie 2017: 55).

case study either. Therefore, we hypothesise that the tendency to use *do*-support will be stronger with lower-frequency verbs than with higher-frequency verbs.

To test this hypothesis, we could not use the lemma frequencies from our sample, since the process of matching *do*-instances to *do*-less instances is likely to have distorted lemma frequencies in the sample: since *do*-instances are generally less common than *do*-less instances, the sample overrepresents verbs favouring *do* and underrepresents verbs avoiding *do*. Instead, for each lexical-verb lemma in our sample we collected its cumulative frequency as attested between 1500 and 1570 in the *Early English Books Online*, abbreviated EEBO (Text Creation Partnership 2014–2016). EEBO is an online collection of more than 146,000 English texts from between 1473 and 1700. An example of a lower-frequency verb, *bargain,* is provided in (142). It occurs only 176 times in all EEBO texts written between 1500 and 1570, whereas higher-frequency verb *usen* in (143) occurs no less than 40,230 times in that same period.

(142)  *Sayth this poore man , ‘ I owe her none , nether <u>dyd</u> I euer* **bargane** *with her for any thinge*

   ‘This poor man says: “I owe her none, nor did I ever bargain with her for anything”.’ (*A Caveat or Warening for Commen Cursetors Vulgarely Called Vagabones*, 1567–1568, PPCEME)

(143)  *And this is the meanyng of that common sentence whiche many* **vse**

   ‘And this is the meaning of that common sentence that many people use.’ (*The Path-way to Knowledg*, 1551, PPCEME)

The cumulative frequencies retrieved from EEBO were included in a regression model (Section 6.4.), where they were compared to the mean. In this model, we investigate whether there exists an interaction effect between source language and lemma frequency.

### 6.3.2.3. Inflection

Section 3.3.2.2. has shown that French loan verbs entering Late Middle English were subject to a markedness bias, meaning that they are more likely to be used in morphologically unmarked forms (e.g. *I* **assume**) than in marked forms (e.g. *he* **assumeth**). Native verbs, in contrast, are not affected by the presence of inflection. We thus expect to find an interaction between inflection and source language, in that the tendency to use *do*-support in French-origin verbs will be even stronger in marked forms than in unmarked forms. For English-origin verbs, however, we do not predict a similar interaction, since inflection on native verbs is not specifically associated with an added processing cost (cf. Section 5.3.4.). Therefore, this chapter will only investigate inflection as an interaction effect, and not as a main effect. It should be noted that we expect the interaction effect between inflection and source language to

remain modest, since we have found that — although inflection is an obstacle to loan word integration — it is not the main one, as the non-finite bias is stronger (cf. Chapter 5.).

In the manual data annotations for this morphological variable we distinguished between unmarked forms, where neither auxiliary *do* nor the lexical verb is inflected, and marked forms, where auxiliary *do* or the lexical verb is inflected. Examples of inflectionally unmarked forms with and without *do* are depicted in (144) and (145) respectively.

(144)   *Then <u>do</u> I **deuide** that corde in the middle , in E*

      'Then I divide that rope in the middle, in E.' (*The Path-way to Knowledg*, 1551, PPCEME)

(145)   *Then from D. to E , I **draw** a right lyne D.H .*

      'Then I draw a straight line D.H. from D. to E.' (*The Path-way to Knowledg*, 1551, PPCEME)

Marked forms include second- and third-person present singular forms, illustrated in (146) and (147) respectively, as well as past forms in all persons and numbers, as in (148).

(146)   *thou **sayest** and thou <u>doest</u> **say***

(147)   *he **writeth** and he <u>doeth</u> **write***

(148)   *we/you/they **spoke** and we/you/they <u>dyd</u> **speak***

For the past, we additionally distinguished cases of the explicitly marked second person, as *dyddest* in example (149).

(149)   *Thou <u>dyddest</u> **promise** once at the sacrament of baptisme to keepe thy fayth*

      'You promised once at the sacrament of baptism to keep your faith.' (*Bishop John Fisher*, 1521, PPCEME)

### 6.3.2.4. Clause structure

The next variable which could influence the use of *do*-support is of a syntactic nature and concerns clause structure. This variable mainly serves as a control variable in the analysis, since the literature has shown that *do*-support is more common with negation and interrogatives than with affirmative declaratives (cf. Huddleston 1976 for Present-day English; Fischer, De Smet & Wurff 2017 for Middle English and Early Modern English). One feature linking negation and interrogatives is that the continuation between subject and verb is typically disrupted, whereas subject-verb order had become dominant over the course of the Middle English period (Fischer,

De Smet & Wurff 2017). Therefore, we hypothesise that Early Modern English clause structures with a non-default SVO word order (e.g. negation and/or inversion) are more likely to host *do*-support than structures with a default SVO word order. More concretely, *do*-support would be more frequent in negative clauses, interrogatives, and negative interrogatives than in affirmative declarative clauses. Example (150) displays an affirmative declarative clause.

(150)   *Then the chefe preste **rose** vp and all they that were with him*

   'Then the chief priest rose up and (so did) all they who were with him.' (*The New Testament*, 1534, PPCEME)

The annotations for negation only included cases with *not*, as in (151), but not adverbs such as *never* or pronouns such as *no one*.

(151)   *I <u>do</u> <u>not</u> **mislike** that you are of conscience so scrupulous*

   'I do not mislike that you have such a scrupulous conscience.' (*The Lyfe of Sir Thomas Moore*, 1534, PPCEME)

The category of interrogatives, next, includes both polar interrogatives (152), which expect a yes-or-no answer, and *wh*-interrogatives (153). However, the regression model did not distinguish between the two types.

(152)   ***Doe*** *ye <u>lacke</u> money?*

   'Do you lack money?' (*Roister Doister*, 1552–1553, PPCEME)

(153)   ***Howe long*** *they <u>stande</u> prating?*

   'How long have they been talking for?' (*Roister Doister*, 1552–1553, PPCEME)

Notice that indirect questions and declarative clauses with rising intonation (characterised by a question mark) were not considered attestations of interrogatives, as both clause types typically have SVO word order.

The last category included in the annotations was the negative interrogative, illustrated in example (154).

(154)   ***Did*** *<u>not</u> you <u>make</u> me a letter brother ?*

   'Did you not write me a letter, brother?' (*Roister Doister*, 1552–1553, PPCEME)

Negative interrogatives typically already imply the polarity of the answer. In example (154), for instance, the expected polarity is positive (i.e. *Yes, I made you a letter*).

Whenever the subject for the above categories was elided, we took the liberty to assume that word order was default and that the continuation between subject and verb was uninterrupted.

### 6.3.2.5. Subject-verb inversion

The fifth variable included in the analysis was subject-verb inversion, a word-order structure in which the verb precedes the subject. That is, the continuation between subject and verb in inversion is disrupted, as was the case for clause structure (Section 6.3.2.4). Whereas subject-verb inversion in Present-day English is restricted to a limited set of contexts[80], it was still relatively common in Early Modern English, and only decreased significantly in the 17th century (Bækken 2000). In the annotations, which indicated the presence or absence of subject-verb inversion, instances of subject elision were considered as following the default word order. An example of subject-verb inversion is displayed in (155), where *cam* precedes subject *the bysshope of Wyncheaster*.

(155)   *The ix day of August* **cam** *the bysshope of Wyncheaster owt of the Towre*

'The ninth day of August the bishop of Winchester came out of the tower.' (*The Diary of Henry Machyn*, 1553–1559, PPCEME)

In Present-day English this example would be translated without inversion, namely as *the bishop of Winchester came*.

Following the NICE-environments described by Huddleston (1976) and the research on word order by Fischer, De Smet & Wurff (2017), we hypothesise that clauses with inversion will be more likely to occur with *do*-support than clauses without inversion. The logic is the same as for interrogatives (see Section 6.2.3.): if the emergence of *do* was in part promoted by a desire to preserve a word order with the subject preceding the lexical verb, inversion contexts should have attracted *do*. Like clause structure (Section 6.3.2.4.), subject-verb inversion is mainly included as a control variable.

### 6.3.2.6. Medial adverbial

Given that the rise of *do*-support has been linked to changes in adverb placement (e.g. Haeberli & Ihsane 2016; Fischer, De Smet & Wurff 2017; cf. Section 6.2.3.), the data annotations also identified cases where one or more adverbials (i.e. a phrase serving as an adverb) are located between the subject and the verb. Such attestations are referred to in this thesis as *medial adverbials*. Medial adverbials can occur both in cases with *do*-support, as in example (156), and without *do*-support, as in example (157).

---

[80] According to the Cambridge Dictionary (2022a), the most common contexts where inversion is used in Present-day English are questions, with negative adverbs, in expressions beginning with *not*, and with *here* and *there*.

(156)   *I do <u>wel</u> **remember** it .*

'I do remember it well.' (*Boethius' Consolation of Philosophy*, 1556, PPCEME)

(157)   *For the beneuolente mynde of a gouernour <u>nat onely</u> **byndeth** the hartes of the people unto hym with the chayne of loue*

'For the benevolent mind of a governor does not only bind people's hearts unto him with the chain of love.' (*The Boke Named the Gouernour*, 1531, PPCEME)

In example (156), for instance, adverbial *wel* is located between the form of *do* (*do*) and lexical verb *remember*. Example (157), next, illustrates how two adverbials (namely *nat* and *onely*) can be placed between the subject (*the beneuolente mynde of a gouernour*) and verb (*byndeth*). One should notice that we annotated cases purely for the presence or absence of medial adverbials (based on their position), and that the number of adverbials between the subject and verb was not accounted for in the analysis. Also, in cases where the subject is elided, we assumed that word order is default.

Since the change of adverb placement (i.e. between the subject and verb) in Middle English could be facilitated by means of *do*-support (cf. Fischer, De Smet & Wurff 2017), we hypothesise that clauses with medial adverbials are more frequent with *do*-support than clauses without medial adverbials. This hypothesis also ties in with the reasoning for clause structure (Section 6.3.2.4) and inversion (Section 6.3.2.5.), namely that *do*-support is more common in clauses where word order is non-default.

### 6.3.2.7. Semantics
The variable of semantics was included as part of a qualitative analysis and was only annotated for cases where *do* occurs when it is not statistically expected (i.e. with English-origin, higher-frequency verbs in affirmative declaratives).

As discussed in Section 6.2.3., *do* in Early Modern English affirmative declaratives developed in analogy to the modals and was essentially an epistemic modal marker of truthfulness, thereby imposing a sense of truthfulness onto the proposition (Budts 2020a, 2020b). Following this reasoning, we hypothesise that *do*-support will be more common in contexts which call for an imposition of truthfulness than in contexts which do not call for an imposition of truthfulness.

### 6.3.3. Data analysis
The annotated sample was analysed in the statistical tool R, using the "lme4" package (Bates et al. 2015) to carry out a mixed-effects logistic regression analysis. The model examined which independent variables significantly affect the dependent variable, namely the presence or absence of *do*-support. The regression model can be analysed

the same way as the models in Chapters 3. and 4. For source language, we additionally plotted a mosaic plot by means of the R-packages "ggmosaic" (Jeppson et al. 2021) and "ggplot2" (Wickham 2016). The quantitative regression analysis including source language, lemma frequency, inflection, clause structure, subject-verb inversion, and medial adverbials was complemented with a qualitative semantic analysis. The latter analysis aimed to gain more insight into those cases where *do*-support is attested, despite it not being predicted by the — wholly non-semantic — variables in the regression model. We additionally calculated a Fisher's exact test to compare the proportions of *do*-support in present versus past forms.

The next section will discuss the quantitative findings for the variables individually: source language in Section 6.4.1., lemma frequency in Section 6.4.2., inflection in Section 6.4.3., clause structure in Section 6.4.4., subject-verb inversion in Section 6.4.5., and medial adverbials in Section 6.4.6. Section 6.4.7. presents the additional qualitative analysis, investigating semantic features of *do*-support.

## 6.4. Findings

This section identifies how the independent variables identified in Section 6.3. impact the dependent variable, namely the distinction between the absence and presence of *do*-support (absence of *do*-support: 0/presence of *do*-support: 1). To this end, we carried out a mixed-effects logistic regression analysis, which predicts correlations between the different variables. On a side note, we verified whether there exist interaction effects between source language and lemma frequency, and between source language and inflection. However, neither interaction was significant. Since interaction effects are known to affect main effects, both interactions were excluded from the model.

The model output of the fixed effects is presented in Table 23, with $p < 0.05$ as significance threshold. The table equally showcases estimates, standard errors, z-values, and p-values given the z-score. The z-values as displayed in Table 23 are rounded to two digits after comma. Notice that the frequencies for the individual lemmas are compared to the mean. For clause structure the reference level is affirmative declarative clauses compared to interrogative clauses, negative interrogative clauses, and negative clauses.

*Table 23: Outcome of mixed-effects logistic regression model for presence or absence of* do-*support (n = 1,592) in PPCEME, sub-period 1500-1569.*

|  | Est. | SE | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | -2.173e-01 | 9.834e-02 | -2.21 | 0.027 |
| Source (French-origin) | 4.028e-01 | 1.226e-01 | 3.29 | 0.001 |
| Frequency | -7.654e-06 | 1.121e-06 | -6.83 | 8.70e-12 |
| Clause (Interrogative) | 2.628e+00 | 4.514e-01 | 5.82 | 5.85e-09 |
| Clause (Negative-interrogative) | 1.707e+00 | 4.521e-01 | 3.78 | 0.000 |
| Clause (Negative) | 9.971e-01 | 2.292e-01 | 4.35 | 1.36e-05 |
| Inversion | 9.011e-01 | 2.149e-01 | 4.19 | 2.76e-05 |
| Medial_adverbial | 1.105e+00 | 1.669e-01 | 6.62 | 3.56e-11 |

Table 23 reveals significant effects for source language, lemma frequency, and clause structure, including inversion and medial adverbials. These effects will now be discussed for each variable separately.

## 6.4.1. Source language

The regression model in Table 23 reveals a positive coefficient estimate (4.028e-01) for source language, which means that source language has a significant effect on the presence or absence of *do*-support (p = 0.001). Concretely, French loan verbs are more common with *do*-support than native English verbs, and this effect is visualised in the mosaic plot below (Figure 13). 39.8% of the attestations in Figure 13 are of French origin, and this percentage coincides with the vertical division between French-origin and English-origin verbs.
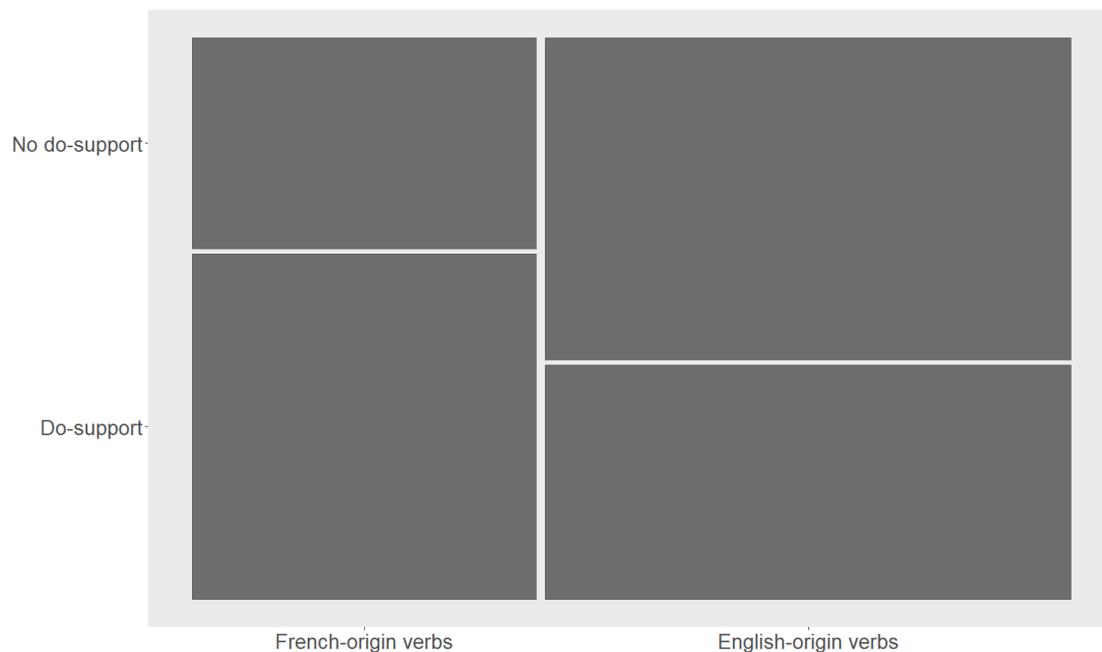
*Figure 13: Distribution of presence and absence of* do-*support for French-origin verbs and English-origin verbs (n = 1,592).*

In other words, as Figure 13 suggests and the statistical analysis confirms, *do*-support is more likely to be used in examples (158) and (159), where the lexical verbs are of French origin, than in examples (160) and (161), where the lexical verbs are of English origin.

(158) *such wines as these <u>do</u>* **trouble** *the head least* .

'Wines as these trouble the head the least.' (*Of Wines and The Vertues of Triacles*, 1568, PPCEME)

(159) *The xxiiij day of September <u>dyd</u>* **pryche** *master doctur Fecknam at Powlles crosse*

'The 24th (day) of September Master Doctor Feckenham preached at Paul's Cross.' (*The Diary of Henry Machyn*, 1553–1559, PPCEME)

(160) *Fyrst they* **saye** *that Martyn luther is a man depely lerned in scryptures* .

'First they say that Martin Luther is a man who is deeply learnt in sculptures.' (*Bishop John Fisher*, 1521, PPCEME)

(161) *Yea* **sayst** *thou me that dame ?*

'Do you say that to me, lady?' (*Roister Doister*, 1552–1553, PPCEME)

This finding corroborates the language contact effect hypothesised in Section 6.3.2.1.

### 6.4.2. Lemma frequency

For lemma frequency, the regression model in Table 23 compares the individual frequencies to the mean. The negative coefficient value (-7.654e-06) in the model suggests that *do*-support is significantly more likely to be selected with lower-frequency verbs than with higher-frequency verbs, the p-value for this effect being 8.70e-12. This means that *do*-support would be more common in low-frequency verb *condemn*, illustrated in example (162), than in high-frequency verb *know*, illustrated in example (163). Whereas *condemn* occurs only 21 times in the entirety of EEBO texts written between 1500 and 1570, *know* occurs 84,911 times.

(162)   *And therefore , we <u>do</u> not **contemne** Rewles , but we gladlie teach Rewles*

    'And therefore, we do not condemn rules, but we gladly teach them.' (*The Scholemaster*, 1563–1568, PPCEME)

(163)   *Our Lord **knowith** my wyll and mynde*

    'Our Lord knows my will and mind.' (*Original Letters, Illustrative of English History*, 1537, PPCEME)

The prevalence of *do*-support in lower-frequency verbs ties in with French loans entering the English language as lower-frequency items, and lower-frequency items — and thus French loans — typically adopting innovative variants (cf. Section 6.3.2.2.). However, an interaction effect between lemma frequency and source language — and thus the effect for lower-frequency verbs being even stronger in verbs of French origin — proves to be only borderline significant (p = 0.06). Because of this, and because interactions in regression models influence the interpretation of the other variables, the interaction was excluded from the model in Table 23.

### 6.4.3. Inflection

Inflection was not included in the model as a main effect, but only as an interaction effect with source language. However, the interaction proved to be insignificant (p = 0.47) and was, therefore, excluded from the model in Table 23. This means that the initial hypothesis is not borne out, and that the tendency to use *do*-support in French-origin verbs is not even stronger in inflectionally marked forms than in unmarked forms. This ties in with the finding that inflection is not the main obstacle to loan word integration, since that role is reserved for finiteness, as found in Chapter 3.

### 6.4.4. Clause structure

The hypothesis for clause structure was that *do*-support would be favoured in non-default word orders — negation, interrogatives, and negative interrogatives —

as compared to affirmative declaratives. The positive coefficients for the three clause types in the regression model reveal that the expectation has been borne out: *do*-support is more frequent in interrogatives (p = 5.85e-09; (165)), negative interrogatives (p < 0.001; (166)), and negation (p = 1.36e-05; (167)) than in affirmative declarative clauses (164), and the effects are strongly significant for all three clause types.

(164)    she **whept** *petefully , and she* **knelyd** *and* **askyd** *God mercy*

'She wept pitifully, and she kneeled and asked God for mercy.' (*The Diary of Henry Machyn*, 1553–1559, PPCEME)

(165)    *Doist thou* **thynke** *it otherwyse ?*

'Do you think otherwise?' (*Boethius' Consolation of Philosophy*, 1556, PPCEME)

(166)    *Do you nott* **know** *my voyce ?*

'Do you not know my voice?' (*Autobiographical Anecdotes of Edward Underhill*, after 1561, PPCEME)

(167)    *He* did not **determin** *to kill the duke of Northumberland*

'He did not determine to kill the duke of Northumberland.' (*The Diary of Edward VI*, 1550–1552, PPCEME)

This means that syntactic structures with non-default SVO word orders are generally more likely to occur with *do*-support than default structures. Following Huddleston (1976) and Fischer, De Smet & Wurff (2017), this effect was expected.

## 6.4.5. Subject-verb inversion

The regression model in Table 23 attests a positive coefficient (9.011e-01) for inversion, one of the control variables in the analysis. This suggests that the presence of subject-verb inversion, as in (168), increases the likelihood that *do*-support will be used (p = 2.76e-05), as opposed to instances where inversion is absent (169).

(168)    *Therfore* doo *I* **set** *one foote of the compas in A , and extend the other vnto D. makyng a part of a circle*

'Therefore, I set one foot of the compass in A., and extent the other to D., making part of a circle.' (*The Path-way to Knowledg*, 1551, PPCEME)

(169)    *So when we* **came** *to Huntyngton they made me to drynke*

'So when we came to Huntington they made me drink.' (*The Autobiography of Thomas Mowntayne*, c1555, PPCEME)

As such, the hypothesis is confirmed, since contexts with non-default word orders are more common with *do*-support than contexts with default word orders, as has also been seen for clause structure.

## 6.4.6. Medial adverbial

As for medial adverbials, the positive coefficient estimate (1.105e+00) in the regression model corroborates that *do*-support is more frequent in clauses with medial adverbials (170) than in clauses without medial adverbials (171).

(170)  I <u>do</u> most humblie and hertelie **thanke** you

'I most humbly and heartily thank you.' (*Letters from Thomas Cromwell*, 1537, PPCEME)

(171)  I **entend** for to gather the summe of this whole matter out of an olde Graecian , named Galen

'I mean to gather the sum of this whole matter out of an old Greek called Galen.' (*Of Wines and The Vertues of Triacles*, 1568, PPCEME)

In other words, contexts with material inserted between the subject and verb increase the likelihood that *do*-support will be selected, and given p = 3.56e-11 this effect is strongly significant. This corroborates the hypothesis, which was based on SVO word order and the changes in adverb placement described in Haeberli & Ihsane (2016) and Fischer, De Smet & Wurff (2017).

Interestingly, in example (170) *do* closely resembles the use of emphatic *do* (Trips 2021 p.c.), conveying an epistemic modal meaning (cf. Budts 2020a, 2020b). This effect of insisting on the truthfulness is reinforced by the presence of medial adverbials *most humblie* and *hertelie*. The function of *do*-support as a marker of truthfulness will be the subject of the next section.

## 6.4.7. Semantics

The previous subsections have predicted in which contexts *do*-support is statistically expected or unexpected. However, the six morphosyntactic variables discussed above cannot account for all cases where *do*-support occurs, since our sample still includes many cases where *do* is attested despite it not being the expected variant. This is especially true for those cases where *do*-support is attested with higher-frequency, English-origin verbs in affirmative declarative clauses without inversion or medial adverbial, as in example (172).

(172)  for all men <u>doo</u> **thynke** that they haue well deserued the same

'For all people think that they have deserved the same.' (*Boethius' Consolation of Philosophy*, 1556, PPCEME)

A closer investigation of the examples reveals that, in both present and past cases where *do* is not expected, as in (172), a semantic motivation is at hand. In example (172), for instance, Boethius uses *do* to highlight the truthfulness of the proposition, namely to stress that all people really think that they deserve the same in life. As such, *do* may be used as a tool to convey truth emphasis, as also seen by Budts (2020a, 2020b) for affirmative *do* in Early Modern English[81]. However, our sample mainly contains instances where *do* conveys adversative (or contrastive) truth emphasis. This principle is illustrated in example (173), which describes how there is still room for benevolence (*do gyue sentence agayne*), even when someone has committed a crime (*one that hath transgressed the lawes*). The contrastive effect is reinforced by the use of *do*-support in *do gyue sentence*.

(173)   *And here it is to be noted , that if a gouernour of a publike weale , iuge , or any other ministre of iustice ,* <u>*do*</u> **gyue** *sentence agayne one that hath transgressed the lawes , or punisseth hym according to the qualities of his trespass , Beneuolence therby is nat any thing perisshed*

‘And here it should be noted that if a governor of public wealth, a judge, or any other minister of justice gives sentence again to someone who has transgressed the law, or punishes him according to the qualities of the trespass, benevolence should not be lost.’ (*The Boke Named the Gouernour*, 1531, PPCEME)

The meaning of *do* has already been shown to convey truth emphasis (cf. Budts 2020a, 2020b), yet the development to adversative truth emphasis is a new finding. A similar development from truth emphasis to adversative truth emphasis has previously been described by Schwenter & Traugott (2000: 11) for adverbial *in fact*. Although they have found three meanings of *in fact*, the focus here is only on *in fact₂*, an adversative adverb with an epistemic meaning which largely resembles the meaning of epistemic adverb *certainly* and adversative adverb *however*. Logically speaking, this use of *in fact* can

signal[…] linguistically that the proposition over which it has scope (q) is more highly ranked on a scale of speaker belief than some proposition (p) that preceded (or is invoked). (Schwenter & Traugott 2000: 13)

As such, this use of *in fact* can convey epistemic belief and truth, and we have now found similar findings for the semantics of *do*. This is confirmed by the fact that *do* is often combined with a marker of (contrastive) truthfulness of some other form, such as an adverbial, a conjunction (*notwithstondinge* in (174)), or a verb (*ensure* in (188)).

---

[81] That *do*-support was used for emphasis has been discussed in some 20th-century sources as well (Rissanen 1985: 177; Koziol 1936 in Denison 1993: 274; Marchand 1938–39, 1939: 123 in Denison 1993: 274).

A striking observation is that — in those cases where *do* is not probabilistically expected — *do*-support is more common in the present tense than in the past tense (Table 24).

*Table 24: Distribution of cases in past and present tense used with and without* do-*support in contexts where* do *is probabilistically avoided (Fisher, p = 0.04).*

|  | With *do*-support | Without *do*-support |
| --- | --- | --- |
| Past tense (n = 194) | 46 (23.7%) | 148 (76.3%) |
| Present tense (n = 294) | 96 (32.6%) | 198 (67.4%) |

Whereas *do*-support is attested in 96 present tense forms (32.6%), it is attested in only 46 past tense forms (23.7%). With a significance threshold at $p < 0.05$, the Fisher's exact test reveals that this effect is just significant ($p = 0.04$). In order to scrutinise this difference between the tenses further, and to provide an accurate description of the semantics of *do* in both cases, the semantic analyses for present (Section 6.4.7.1.) and past tense cases (Section 6.4.7.2.) will be conducted separately.

### 6.4.7.1. Present tense

An example of a present tense proposition conveying adversative truth emphasis is presented below. (174) depicts part of the Exodus and describes how God gave Moses' people good land (*it is a good lande which the Lorde oure God doeth geue us*), although they were disobedient (*Notwithstondinge ye wolde not consente to goo vpp , but were dishobedient vnto the mouth of the Lorde youre God*).

(174)  *And they departed and went vp in to the hye contre and came vnto the ryuer Escoll , and serched it out , and toke of the frute of the londe in their hondes and brought it doune vnto us and brought us worde agayne and sayde : it is a good lande which the Lorde oure God* <u>*doeth*</u> **geue** *us . Notwithstondinge ye wolde not consente to goo vpp , but were dishobedient vnto the mouth of the Lorde youre God , a~d murmured in youre tentes and sayde : because the Lorde hateth us , therfore he hath brought us out of the londe of Egipte , to delyuer us in to the handes of the Amorites and to destroye us .*

‘And they departed and went up into the high country and came to the river Escol, and searched it and took the fruit of the land in their hands and brought it down to us and brought us word again and said: "It is a good country which the lord our God give us". Notwithstanding you did not consent to go up, but were disobedient to the word of the Lord, your God, and murmured in your tents and said: "It is because the Lord hates us that he has taken us out of the

country of Egypt, to deliver us to the hands of the Amorites and to destroy us".' (*William Tyndale's Five Books of Moses Called the Pentateuch*, 1530, PPCEME)

Apart from *do*, the contrast between the two propositions in example (174) is further reinforced by the use of contrastive conjunction *notwithstondinge*.

Another example of present tense adversative truth emphasis is provided in (175), which states that, although we are not consciously aware of it (*without thynkyng thereon howe it is digested* and *not knowyng thereof*), yet we digest our food and breathe in our sleep (*we do take wynde and breathe in slepe*).

(175)   *And I do not speke nowe of the voluntarye mouynges of the soule , that hath knowledge , but of the natural intencion of thynges , euen as it is that we do digest meates , that we haue eten without thynkyng thereon howe it is digested and as we <u>do</u> **take** wynde and breathe in slepe , not knowyng thereof .*

   'And I do not speak now of the voluntary motions of the soul, which have knowledge, but of the natural intention of things, even as it is that we digest food, which we have eaten without thinking about how that it is digested and, as we breathe in our sleep, not being aware of it.' (*Boethius' Consolation of Philosophy*, 1556, PPCEME)

The two facts in example (175) are being opposed by the use of *even as* complemented with *do*.

A last example of adversative emphasis in the present tense is displayed in (176), where the treacle in question is generally not recommended for frequent use (*no body to take this triacle in the heate of Sommer . Neyther ought it to be taken oft and much*), but its frequent use is recommended for older people (*I counsell also that they whose yeres turne towards age doe take it oft and much*). Apart from *do*-support, this contrast is reinforced by *also*.

(176)   *But I would counsell no body to take this triacle in the heate of Sommer . Neyther ought it to be taken oft and much , of them that are not of flourishing or lusty age , neither of the~ that are of hote natures or complexions : I counsell also that they whose yeres turne towards age <u>doe</u> **take** it oft and much , not with water but with wine .*

   'But I would counsel nobody to take this treacle in the heat of summer. Neither should it be taken often nor much by them who are not pregnant or of young age, neither by them who are of warm natures and complexions. I also counsel that they whose years turn towards age take if often and much, not with water but with wine.' (*A Book of Wines, by William Turner*, 1568, PPCEME)

In line with the regression model outcome (Section 6.4.), our sample also contains attestations where *do*-support is not expected and, therefore, not used. Interestingly,

those attestations mainly occur in complement-taking predicates and backgrounded clauses (often constituted of cognitive and perception verbs), examples of which are listed in (177), (178), and (179) below.

(177) *Thou* **seest** *how that faythe helped his workes and how of the workes his faythe was made perfyt .*

'You see how faith helped his works and how, because of the works, his faith was made perfect.' (*Bishop John Fisher*, 1521, PPCEME)

(178) *Thou then as I* **thynke** *didest deceyue Fortune with thy gloryous wordes , when that she thus fauored the , and cheryshed the as her owne derlynges .*

'You then, as I think, deceived Fortune with your glorious words, when she thus favoured you, and cherished you as her own darling.' (*Boethius' Consolation of Philosophy*, 1556, PPCEME)

(179) *And two causes I* **finde** *why he was ordeyned there*

'And I find two causes as to why I was put in order there.' (*The Anatomie of the Bodie of Man, by Thomas Vicary*, 1548, PPCEME)

In example (177), for instance, *thou seest* is background information which can easily be excluded. The same principle applies to example (178), where *as I thynke* could be excluded without changing the contents of the sentence, namely that the interlocutor deceived Fortune. In example (179), last, *two causes I finde* could easily be replaced by *there are two causes*, undermining the importance of *I finde*.

Other contexts which tend to remain *do*-less are hypothetical clauses and mental-space builders, as illustrated in examples (180), (181), and (182).

(180) *the other is called uenacelis , of whom it is too much to treate of now , vntyll we* **come** *to the anathomie of the wombe , &c.*

'The other is called vena cava, which is too much to treat right now, until we come to the anatomy of the womb, etc.' (*The Anatomie of the Bodie of Man, by Thomas Vicary*, 1548, PPCEME)

(181) *By gogs bones when he* **cometh***, now that I know the matter He shal sure at the first skip , to leape in scalding water*

'By Gog's [i.e. God's] bones when he comes, now that I know the matter he shall surely at the first skip leap in the scalding water.' (*Gammar Gvrtons Nedle*, c1553–1563, PPCEME)

(182)  *And if any sorance come to an oxe , $and $he waxe olde , broysed , or blinde , for ii. s. he maye be fedde , and thanne he is mannes meate , and as good or better than euer he was . And the horse , whan he* **dyethe***, is but caryen .*

'And if an ox has any type of sore, and he becomes old, bruised, or blind, for two days he may be fed, and then he is man's meat and as good as or better than he ever was. The horse, in contrast, is just carried to burial when he dies.' (*The Book of Husbandry, by Master Fitzherbert*, 1534, PPCEME)

In example (180), *vntyll we* **come** *to the anathomie of the wombe* is used as a mental-space builder, referring to one of the next sections in the anatomy handbook. Examples (181) and (182), next, illustrate the use of hypotheticals, as *when he* **cometh** ('in case he comes') and *whan [the horse]* **dyethe** ('in case the horse dies') correspond to *He shal sure at the first skip , to leape in scalding water* and *is but caryen* respectively.

A last observation in the qualitative semantic analysis for present tense cases is that syntactically backgrounded information, for instance in relative clauses, is often used without *do*-support. This phenomenon is illustrated by the examples in (183), (184), and (185).

(183)  *Of the Natures , properties , profits , hurtes and helps that* **come** *of Wyne .*

'of the nature, properties, profits, hurts, and helps that come of wine' (*A Book of Wines, by William Turner*, 1568, PPCEME)

(184)  *Therefore if I parte that line B.D , in the middle in to two equall portions , that middle pricke whiche here is F is the verye centre of the sayde circle that I* **seke** *.*

'Therefore, if I part that line B.D. in the middle to two equal portions, that middle point which here is F. is the very centre of the said circle which I seek.' (*The Path-way to Knowledg*, 1551, PPCEME)

(185)  *For it lyeth in your owne power , what fortune you had leuer haue , that is to sai : to take what fortune ye wyll . For all fortune that* **semeth** *sharpe or euyll yf it do not exercyse the good folke , or correct and chastyce the wicked folke it greueth or ponysheth .*

'For if it lies in your own power, what fortune would you rather have had? That is to say, take what fortune you want. For all fortune which seems violent or evil causes difficulty or punishment if you do not educate the good people, or correct and chastise the evil people.' (*Boethius' Consolation of Philosophy*, 1556, PPCEME)

The attestations in the three relative clauses above (*that come*, *that I seke*, and *that semeth*) are all used without *do*-support.

### 6.4.7.2. Past tense

The semantic analysis of past tense cases with *do*-support reveals a similar picture to the one painted for present tense cases, namely that *do* is often used to convey adversative truth emphasis. In example (186), for instance, *do* occurs when an uncertain situation in the narrative finally comes to a resolution, namely when Sir Thomas Arundel was cast of felony in treason.

(186)  *Sir Thomas Arrundel was likewise cast of felony in treason , after long controversie ; for the matter was brought in trial bie seven of the cloke in the morning 28. day ; at none the quest went togither ; they sate shut up togither in a house , without meat or drinke , bicause they could not agree , al that day and all night ; this 29. day in the morning they <u>did</u> **cast** him .*

'Sir Thomas Arundel was likewise cast of felony in treason after long controversy, for the matter was brought in trial by seven o'clock in the morning of the 28th. At noon the trial came together. They sat in a house together, silent and without food or drink, because they could not agree all day and all night long. The 29th in the morning they cast him.' (*The Diary of Edward VI*, 1550–1552, PPCEME)

*Do* conveys implicit contrast between the past situation (*they sate shut up togither in a house*, *they could not agree*) and what is happening at the moment of action, when *they [finally] did cast him*.

Contrast is also conveyed in example (187), which opposes the fact that subject *he* ate, and the fact that he only ate very little. The verb *eten* ('to eat') is used with *do* to reinforce the truthfulness of the proposition.

(187)  *And albeit he <u>did</u> **ete** but verey moderatly ; yet after our retorne home not oonly his flux began to encreace vpon hym , but also the feuer toke hym farvently . Wherupon Phisicions were callyd for help , who after they perceyved the fever to bee contynuall without intermission and the flux to encrease to a voyding of blud , mynestred vnto hym suche medicynes as they thought moost convenyent*

'And although he ate, he ate very moderately. Yet after our return home not only his flux began to increase, but also fever took him fervently. Upon this matter doctors were called for help, who — after they perceived the fever to be continual without intermission and the flux to increase to a bloodletting — served him the medicines which they deemed most appropriate.' (*Original Letters, Illustrative of English History*, 1525, PPCEME)

Example (188), finally, depicts the use of past tense *do*-support for truth emphasis with absence of contrast, as has also been described by Budts (2020a, 2020b).

(188)  *Madame , I recommend me unto you , doyng you to undrestonde that I have receyved your Lettres by your servante concernyng the maryage of your doughter , by the whiche I do perceyve that the Gentilwoman beyng accompaned with your said doughter unto your howse , hath informed you that it was my mynde for hir to certyfye you that the Controwler of the Pryncesse howsolde dothe bere hys synguler favour to your said doughter. Trewly she mysusyd hir selff in gevyng you any suche knowlege on my behalff , for I ensure you that I <u>dyd</u> **geve** unto hir no comaundment so to do , for at that tyme I had harde no comunycacion touchyng that matter .*

‘Madam, I am writing you, letting you understand that I have received the letters by your servant concerning the marriage of your daughter, by which I perceive that the lady being accompanied to your house with your daughter has informed you that it was my mind for her to certify you that the controller of the princess bears his singular favour to your daughter. Truly she misused herself in giving you any such knowledge on my behalf, for I ensure you that I did not give her any command to do so, for at that time I had no communication touching upon that matter.’ (*Original Letters, Illustrative of English History*, 1538, PPCEME)

In the example, the author of the letter reinforces the fact that they did not give the receiver’s daughter any command to convey the message by using the verb *ensure*, and a form of *do*-support with the verb *give*.

As seen from the Fisher’s exact test in Section 6.4.7., past tense cases are less frequent with *do*-support than present tense cases, and this effect is just significant ($p = 0.04$). A possible explanation is that truth emphasis in the past tense is more redundant than in the present, as the past — and the truth as it was in the past — tends to be known by the time the utterance is being made.

## 6.5. Discussion

This chapter has dealt with the strong rise of *do*-support in 16th-century English. Whereas *do* was originally used as an auxiliary verb, it later evolved into a semantically empty operator to form emphatic, interrogative, inverted, and negative clauses. Diachronic evidence has shown that, while *do*-support is generally not used in affirmative declaratives in Present-day English anymore, this is the context in which it actually emerged in Early Modern English. Based on clues from the literature as well as our previous findings for the non-finite bias, we have hypothesised that the high borrowing rates from French in 14th-century English and the greater usage of *do*-support in affirmative clauses in 16th-century English are connected (Section 6.2.4.). That is, verbs of French origin, which are subject to a non-finite bias (cf. Chapter 3.), may also have been more frequent with *do*-support than verbs of

English origin. Since French loans typically entered the language as low-frequency items adopting newer syntactic variants, and the rates of French loans were so high, French influx may have raised the general rates of verbs being slotted in *do*-support. In Late Middle English this may have happened by promoting the grammaticalisation and usage of *do*-support, and in Early Modern English by boosting the reliance on the already existing construction of *do*-support. This hypothesis has been tested in a corpus study on PPCEME data, for which the data and methodology have been described in Section 6.3. Notice that — apart from source language — we have also investigated the effect of six other variables on the use of *do*-support: lemma frequency, inflection, clause structure, subject-verb inversion, the presence of medial adverbials, and the role of semantics. Register, style, and genre were controlled for by using a text-matched sample of attestations with and without *do*-support. The quantitative findings (including all variables but semantics) in Section 6.4. have suggested three major contexts in which *do*-support in Early Modern English is significantly overrepresented. First, *do*-support is significantly overrepresented in non-default SVO word orders, including interrogatives and negatives, but also when material is inserted between the subject and verb. This is consistent with the literature on *do*-support (cf. Huddleston 1976; Fischer, De Smet & Wurff 2017). However, with all this controlled for, *do*-support is also strongly favoured with French loan verbs as compared to native English verbs. Although such an effect of source language on the rise of *do*-support has been proposed in the literature (e.g. Fischer & Wurff 2006: 155; Moretti 2021, subm.: 25), it has — to our knowledge — never been systematically investigated before. In addition, it is consistent with typological research stating that *do*-support-like constructions can be used as a strategy for loan verb accommodation (e.g. Jäger 2004: 10). The reason for this bias for loan verbs is presumably the non-finite bias previously found for French loan verbs (cf. Chapter 3.), and *do*-support acting as a periphrastic tool to use the loan verb non-finitely. Third, *do*-support is more common with lower-frequency than with higher-frequency verbs, and this is especially true for French-origin verbs. This finding is not surprising since lower-frequency items tend to be innovative and hence adopt newer patterns and constructions, such as the — at the time — developing construction of *do*-support. All three contexts where *do*-support is the preferred option — non-default syntactic word orders, French loan verbs, and lower-frequency items — have in common that they come with a higher processing cost than native verbs, higher-frequency items, and default syntactic word orders respectively. For loan words and low-frequency items extensive evidence for the increased cost was provided in Sections 5.3.2. and 5.3.4.; for clauses with non-default SVO word orders (e.g. negatives, interrogatives) the elevated processing cost comes from their lower frequency compared to default SVO word order (e.g. affirmative declaratives). In all three cases, which put a burden on processing, *do*-support can fulfil a facilitator function by allowing for the

periphrastic use of the lexical verb and by compensating for the higher processing cost.

However, some attestations of *do*-support have also been encountered — albeit to a much lesser extent — with English-origin, higher-frequency verbs in affirmative declarative clauses. A qualitative study has revealed that *do*-support in such cases, where it is likely to be avoided due to morphosyntactic reasons, has a semantic motivation (Section 6.4.7.): it tends to convey both truth emphasis and adversative (or contrastive) truth emphasis. This aligns with previous findings by Budts (2020a, 2020b) that *do* in Early Modern English affirmative declaratives was used as an epistemic modal marker underlining the truthfulness of a proposition. This also explains why *do*-support occurs significantly more often in the present tense than in the past tense, namely because truth emphasis in the past tense is not as necessary, since the truth as it was in the past is generally known at the moment of the utterance.

The findings in this chapter have suggested that *do*-support in 16th-century English was used as a 'jack-of-all-trades', fulfilling various functions. The empiric evidence has corroborated the language contact hypothesis, which means that French-origin verbs were more common with *do*-support than English-origin verbs. As such, the rise of *do*-support — and possibly also of other newer constructions — may reflect an accommodation strategy for French loan verbs in English. This is preliminary evidence of loan word accommodation biases *indirectly* having prolonged impact on their recipient-language grammar. Then again, we do not claim that the rise of *do* was solely due to language contact between Middle English and French: on the contrary, this chapter has shown that *do*-support has not emerged in a vacuum, and that a multiplicity of factors is potentially involved in the variation between presence and absence of the construction. *Do*-support in Late Middle English was likely already developing and its use was already increasing due to various factors, both language-internal and -external (e.g. contact with French in Denison 1993; word order changes in Warner 1993 and Fischer, De Smet & Wurff 2017)[82]. Although it is unlikely that the integration and accommodation of French loan verbs in English has caused the rise of *do*-support, loan verbs may have played a role in accelerating the already ongoing rise. This ties in with Thomason's (2001) idea that, although "changes are actually motivated by internal pressure within the language", "they would have been less likely to happen if the initial contact-induced change had not happened".

A question resulting from the findings is why the rise of *do*-support has only started in the Late Middle English period, whereas French had already started influencing the English language several hundreds of years before. Moreover, by the time French loan

---

[82] Although traditionally there exists a dichotomy between language-internal and -external motivations (as seen in Farrar & Jones 2002: 1), this research is an example case of both types of motivations interacting.

verbs were used with the facilitative strategy of *do*-support, the language had already lost many of its (verbal) inflections, and the need for a facilitative strategy had decreased[83]. This ties in with the actuation problem as formulated by Weinreich, Labov & Herzog (1968: 102): "[w]hy do changes in a structural feature take place in a particular language at a given time, but not in other languages with the same feature, or in the same language at other times?" Whereas the actuation problem has often been deemed unsolvable, in this case the problem — namely why the rise of *do*-support did not develop "in the same language at other times [i.e. earlier]" — may actually be solvable. As described in Section 6.2.3., Early Modern English *do* was influenced by analogical pressure from the modal verbs, which developed into a separate auxiliary category. Similar to the modals, *do* in *do*-support marked both tense and mood, and it developed an epistemic modal meaning of truthfulness (Budts 2020a, 2020b). According to Warner (1993: 235), "the rise of periphrastic *do* is closely associated with the development of basic status for auxiliaries". Additional evidence of this claim stems from the fact that pro-verbal *do* cannot be combined with other auxiliaries or modals (Warner 1993: 152). Following Warner (1993: 223) and Budts (2020a, 2020b), the development of Early Modern English *do* happened in analogy to the development of the modal verbs, which had grammaticalised in the 12th century (Denison 1993; Warner 1993: Ch. 8.), and which hence paved the way for the rise of *do*-support. However, this did not happen until the beginning of the Early Modern English period, since the modals continued to develop until then.

A caveat of this chapter is that we have not included word length as a variable in the analysis. This may correlate with the variable of source language and play a role in the variation between the presence and absence of *do*-support as well, since French-origin verbs typically contain more syllables than inherited English-origin verbs (compare *countremaunden*, *performen*, and *remembren* to *bidden*, *hurten*, and *seien*). The number of syllables relates to prosody, and prosody has been said to affect syntactic structure, such as word order (cf. Elfner 2018: 4). Another factor which may be interesting to investigate is the age of the language user: Warner (1993: 237–239) has found evidence for the use of *do* being inversely proportional to the language user's age, as its use decreases with age. A last interesting avenue for future research would be to investigate whether the shift from affirmative to emphatic *do* may have happened through medial adverbials. That is because *do* in example (170) closely resembles the use of emphatic *do*. In light of this, it may be valuable to categorise the medial adverbials semantically into modal and non-modal types.

This chapter has provided evidence for loan word accommodation biases having long-term consequences on the syntax of their recipient language. At a more general

---

[83] That said, verbal inflections were not lost in all finite categories. The present second and third persons singular, for instance, remained to be inflected.

level, language users of languages exposed to a high influx of loan words may prefer periphrastic strategies, either existing or new ones. The findings particularly apply to the French-Middle English contact setting, where the accommodation biases of French loan verbs may have reinforced the drift of English towards non-finiteness, such as the rise of *do*-support. Since *do*-support still exists in Present-day English and in some North-Germanic languages, yet it does not exist in other West-Germanic languages such as German and Dutch, this chapter has shown that French has *indirectly* contributed to some of the most striking divergences between English and the continental West-Germanic languages. This raises the question: was the impact of French influx on the syntax of English restricted to *do*-support, or did it equally affect other constructions? A logical follow-up study is to investigate one of the other constructions emerging in Late Middle English. Therefore, the next chapter will present a smaller case study on the rise of light verbs at the end of the Middle English period.

# Chapter 7

# The rise of light verbs in Late Middle English

## 7.1. Introduction

The previous chapter has dealt with the possible long-term impact of loan word accommodation biases on recipient-language grammars, more specifically in the case of French loan verbs entering Middle English. The findings have shown how the already ongoing rise of *do*-support in Early Modern English may have been accelerated by the high borrowing rates from French in Late Middle English. This *indirect* effect of language contact may add to linguistic theories explaining why English diverges from the continental West-Germanic languages with regard to the presence and use of *do*-support. Moreover, these findings raise the question whether the impact of French loan words on the English language is limited to the rise of *do*-support, or whether it may also indirectly have triggered other syntactic developments. For example, simultaneously with *do*-support, the English language developed or came to increasingly rely on a set of periphrastic constructions, such as modal verbs, infinitives, gerunds, progressive constructions, and light verbs (e.g. *make* in *to* **make a** *decision*). The use of *do* in *do*-support resembles the light verb strategy from a syntactic angle[84], since in both structures the grammatical information is carried by *do* or the light verb. Therefore, the focus of this chapter is on the light verb construction, which constitutes an essential part of the English language. This chapter will present a smaller-scale case study on the potential prolonged impact of French influx on Late Middle English (1350–1500).

Similar to *do*-support, the rise of light verbs in the later period of Middle English unfolds under a linguistic climate of rapid change (Denison 2003: 68; as proven quantitatively by Nevalainen et al. 2020: 26). For instance, Late Middle English is characterised by an overall loss of inflectional morphology, which includes the reduction of verbal inflections (e.g. Pyles & Algeo 1982: 153–154; Rissanen 2000). A related development was the increasing reliance on a number of recently emerging

---

[84] The light verb strategy has also been referred to as the '*do*-strategy' (Wohlgemuth 2009: 104).

periphrastic (non-finite) structures, including light verb constructions (e.g. Görlach 2003: 97; Green 2017). A more detailed account of the linguistic context of this period can be consulted in Section 6.1.

In what follows we will first define light verb constructions as a type of complex predicate (Section 7.2.1.). Then we will discuss the two main functions of light verbs (Section 7.2.2.), and we will expand on their usage in English (Section 7.2.3.), which emerged in the Late Middle English period. Section 7.2.4. briefly elaborates on the use of light verb constructions in French, showing that light verbs are frequent in French as well, and that French and English are (and have always been) compatible in this. As seen in Chapter 5., loan words tend to be favoured with periphrastic strategies, where they carry less grammatical information, and light verb constructions allow for a periphrastic use of the lexical verb. Furthermore, nouns are more borrowable than verbs (cf. Section 7.2.2.), and lexical verbs combined with light verbs can be used nominally. Therefore, Section 7.2.5.1. will hypothesise that — like *do*-support — Middle English light verb constructions are characterised by an overrepresentation of French loan words compared to native English words. This is in line with a study by Ronan (2012, 2014), who has found the same overrepresentation in Chaucer's *The Canterbury Tales*. As such, the steep rise of light verbs may have been promoted by the existence of accommodation biases in French loan verbs in Middle English. In Section 7.2.5.2., we will hypothesise that English-origin deverbal nouns in light verb constructions are more often modified than French-origin nouns. The reasoning is that French-origin nouns are mainly used in light verb constructions to ease loan verb accommodation (cf. language contact hypothesis), and their use will hence be less motivated by other functions, such as modifying the verbal expression. Both hypotheses will be tested by means of a small-scale case study using data from the Helsinki Corpus. The data and methodology will be presented in Section 7.3., where special attention will be devoted to the tests which were applied for identifying constructions as light verb constructions. In Section 7.4., we will first present the findings for the language contact hypothesis. Each of the eight light verbs under investigation are discussed separately: *bear* (Section 7.4.1.1.), *do* (Section 7.4.1.2.), *give* (Section 7.4.1.3.), *have* (Section 7.4.1.4.), *make* (Section 7.4.1.5.), *put* (Section 7.4.1.6.), *set* (Section 7.4.1.7.), and *take* (Section 7.4.1.8.). Our findings confirm that the nouns used in light verb constructions are predominantly of French origin, which indicates that the rise of light verbs in Late Middle English may have been boosted by French influx. However, the modification hypothesis will not be corroborated by our data (Section 7.4.2.). An in-depth discussion of the findings as well as their implications conclude this chapter in Section 7.5.

## 7.2. Light verb constructions

### 7.2.1. Definition

Light verb constructions are a type of *complex predicates* (cf. Wohlgemuth 2009: 102), or "'multiword' sequences acting like a single predicate" (Nash & Samvelian 2016: 1). An example of a complex predicate is the use of a form of *faire* ('to do, make') + infinitive in Modern French, as in example (189).

(189)   *La maman* **fait manger** *l'enfant.*

       'The mother lets (lit. does) the child eat.' (Bezinska & Novakova 2010: §13)

In complex predicates, "[e]ach member of the sequence contributes to the argument structure and the semantic roles assigned by the predicate" (Nash & Samvelian 2016: 1). This type of construction is found across languages in the world (Anderson 2011: 795), and the literature on complex predicates is rich (e.g. Spencer 1991; Ackerman & Webelhuth 1997; Alsina, Bresnan & Sells 1997), but also particularly intricate, with many different terms and definitions adopted for slightly different constructions (Bowern 2008: 162; Anderson 2011: 795). This chapter focuses on one subcategory of complex predicates (cf. Anderson 2011: 796; Los et al. 2012: 7, 18), namely light verb constructions, such as *to take notice of.*

Like complex predicates, light verb constructions, commonly abbreviated as LVCs, have been referred to in various ways: 'expanded predicates' (Algeo 1995: 203–204), 'stretched verb constructions' (Allerton 2002: 6), 'thin verbs' (Mel'čuk 1998), 'verbo-nominal combinations' (Claridge 2000: 69), etc. (for an overview, see Iglesias-Rábade 2001: 143). However, these may refer to slightly different constructions. According to Anderson (2011: 81), the term 'light verb construction' is generally overused and should be restricted to the 'true' sense of the term, namely

> a complex predicate subtype in which there are two components, consisting of a lexical stem (belonging to some part of speech category […]) and an inflectable verbal element, the 'light' verb.

This will be the definition adopted in this thesis. Simply put, light verb constructions are lexicalised combinations of a light verb and a deverbal noun or nominal (also called action noun or predicate noun). The light verb (e.g. *do*, *give*, *have*, *take* or *make*) is 'light'[85] in a semantic sense (Cattell 1984: 2), since it is a "general-purpose verb" used to express an action or activity (Clark 1993: 29–30)[86]. Hence, it is the deverbal noun which carries the lexical meaning, whereas the light verb is inflected and carries

---

[85] According to Cattell (1984: 2), the term 'light' in this sense was first coined by Otto Jespersen.
[86] Elenbaas (2013: 49) stresses that, "[a]lthough light verbs do not express a full lexical meaning, they are not entirely meaningless".

the grammatical information (Cattell 1984: 2). Note that light verbs tend to be of English origin, and that the deverbal noun is sometimes followed by a prepositional phrase (e.g. *to have knowledge **of***). Light verb constructions can be contrasted with *simplex* (also called simple, single-word, or full) *verbs*, which are verbs consisting of a single word (e.g. *to know*).

Like many other languages[87] (see e.g. references in Mohanan 2006; Wohlgemuth 2009: 102–103), English has various light verb constructions in the true sense of the term (cf. Anderson 2011: 81). Four examples of Middle English light verb constructions are displayed in (190), (191), (192), and (193). While examples (190) and (191) are instances of *have* as a light verb, (192) illustrates the use of *take*, and (193) of *do*.

(190) *And we honore anely þat thing Wharof we **haue** kind **knawing**.*

'And we honour only that thing that we know innately (lit. have innate knowing of).' (*The Northern Homily Cycle*, c1315, HC)

(191) *Of alle men he wuld **haue doute***

'He would doubt (lit. have doubt of) everyone.' (*Handlyng Synne*, 1303, HC)

(192) *men moten **taken heede**, what is seid of Crist bi his godheed, and what bi his manheed*

'Men must beware (lit. take heed) what is said of Christ by his godhood, and what by his manhood.' (*Purvey's General Prologue to the Bible*, a1450, HC)

(193) *yt was fully assented ferst to **haue don execucion** vpon any man that had be proued giltyf*

'All consented to first execute (lit. do execution on) any man that had proven guilty.' (*The Appeal of Thomas Usk against John Northampton*, 1384, HC)

In example (190), the light verb construction is *haven knouing* (lit. 'to have knowledge', meaning 'to know'): the author opts for the English-origin noun *knouing* instead of the verb *knouen*. Example (191) exemplifies light verb construction *haven doute* (lit. 'to have doubt', meaning 'to doubt'), based on Old French *do(u)ter*. Instead of the verbal form *douten*, the loan is used in its associated nominal form, namely *doute*. Example (192), next, shows the construction *taken hed* (lit. 'to take heed', meaning 'to heed'), where the English-origin noun *hed* is used rather than the verb *heden*. The last example, (193), illustrates how the French-origin noun *execucioun* is used in the light

---

[87] Anderson (2011: 795) even goes as far as saying that "[c]omplex predicates of one sort or another are found in virtually all languages".

verb construction *don execucioun* (lit. 'to do execution', meaning 'to execute'). Instead of the verb *executen*, the author opts for the associated noun, *execucioun*.

Even if there exists a corresponding verb, light verb constructions are typically used with the associated deverbal noun, or the noun derived from the verb or verb phrase. Whereas the noun in some constructions may "correspond to a verb of similar meaning" (e.g. *to give an answer* – *to answer*; *to take a walk* – *to walk*), in other cases it does not (Algeo 2006: 269). This could be because there exists no corresponding simplex verb (e.g. *to do homework* – *\*to homework*; *to make peace* – *\*to peace*), because the noun and the semantically equivalent verb are not cognates (e.g. *to take cover* – *to hide*; *to do a favour* – *to help*), or because the cognate simplex verb is not a semantic equivalent of the noun (e.g. *to take a chance* – *to chance*) (Algeo 2006: 269). The light verb which is most commonly combined with deverbal nouns is *to do* (Wohlgemuth 2009: 104).

Stylistically speaking, light verb constructions tend to be linked to a higher degree of colloquialness or informality than simplex verbs. Quirk, Greenbaum & Leech (1985: 751; also see Curme 1983: 22), for example, have stated that a light verb construction

> provides greater weight than the corresponding SV type [i.e. simplex verb], especially if there are no optional adverbials, and is often preferred to the SV construction in informal language.

However, the idea that light verb constructions are more informal than simplex verbs has also been challenged by other linguists (e.g. Brinton 1996: 190; for an overview of authors, see Shahrokny-Prehn & Höche 2011: §3).

## 7.2.2. Functions

Light verb constructions can be used to enable internal modification of the deverbal noun (Nickel 1968: 15; Akimoto & Brinton 1999: 51–53; Bonial 2014), as in example (194), where *wonder* ('wonder') is modified by the qualifying adjective *grete* ('great').

(194)  *Þai* **had** *grete* **wonder**

> 'They were very curious (lit. They had great wonder).' (*The Northern Homily Cycle*, c1315, HC)

Bonial (2014: 8, 93) even calls nominal modification the "primary function" of light verb constructions, and argues that

> it is so common for some type of modification to exist within LVCs (adjectival modification, relativization of the noun object) that it is widely noted in the literature on LVCs.

Indeed, light verb constructions are characterised by their "flexibility of verbal modification" (Brinton & Akimoto 1999: 2) compared to simplex verbs.

Apart from modification purposes, light verbs are cross-linguistically also resorted to for loan verb integration (cf. Wohlgemuth 2009), as was discussed in the state of the art (Section 2.3.1.3.). In the light verb strategy, deverbal loan nouns are combined with a light verb,

> which has an auxiliary-like function and bears the inflection or — more generally — all grammatical information of the resulting compound predicate, while the semantic information is by and large associated with the loanword part of the complex verb. (Wohlgemuth 2009: 102)

This strategy is the second most frequent loan verb accommodation strategy identified by Wohlgemuth (2009), and it is used, for instance, to integrate Chinese loan verbs in Japanese (e.g. Miyamoto 1999). The high frequency of the light verb strategy ties in with its rather low integrational effort needed to accommodate loan verbs (Wohlgemuth 2009: 136). It can also be linked to the fact that loan words can be used as nouns, which are at the top of the hierarchy of borrowability and which are, therefore, more borrowable than verbs (e.g. Muysken 1981; Poplack, Sankoff & Miller 1988; Matras 2007, 2009; Winford 2010; Hock & Joseph 2019). Loan noun integration generally consists of assigning number and grammatical gender to the noun (e.g. Scherling 2013: 46). Not only do nouns carry fewer inflections than verbs (Harris & Campbell 1995), they are also less conceptually complex (Meillet 1921; Matras 2007, 2009). Additionally, nominal inflections are borrowed more frequently than verbal inflections (Seifart 2017: 424). According to Poplack, Sankoff & Miller (1988: 62), that is because verbs have more highly bound morphemes, and "the more highly bound the morpheme (such as inflections, case endings, function words), the less likely it will be to undergo borrowing". Hence, loan verbs often enter their recipient language as another part of speech, for example as nouns, which can then gradually evolve into verbs (Whitney 1881). Indeed, Hock & Joseph (2019: 227) have found that

> if the need for borrowing does arise, many languages instead borrow a nominal form of the verb and employ a native all-purpose verb such as *do* or *make* as a means of turning that form into the equivalent of a verb.

Language users have even been found to code-switch in the middle of light verb constructions, for example between Spanish and English (Balam, Parafita Couto & Stadthagen-González 2020), but this finding will not be developed here. Since the light verb strategy as a tool for loan verb accommodation was already extensively discussed in the state of the art (Section 2.3.1.3.), it will not be expanded on any further in this chapter.

Other functions of light verb constructions, such as pragmatic reordering and meaning specification, are discussed in Ronan (2012: 147–148).

### 7.2.3. Light verb constructions in English

Akimoto & Brinton (1999; also see Ronan 2012: 150–151, 2014: 16) were the first linguists to conduct actual corpus research on the use of light verbs in Old English. They have shown that the first light verb constructions had already developed in Old English, yet "the Old English 'V + N' collocation is less fully grammaticalized, idiomaticized, and lexicalized than its Modern English counterpart" (Akimoto & Brinton 1999: 54). Old English light verb constructions combined *ge(don)* 'do', *(ge)macian* 'make', *sellan* 'sell', *giefan* 'give', *niman* 'take', or *habban* 'have' with a deverbal noun (Akimoto & Brinton 1999: 23). "[R]ivalry [existed] between *don* and *macian*, between *sellan* and *giefan*, and between *niman* and *tacan*" (Akimoto & Brinton 1999: 23). Two instances of Old English light verb constructions, cited from Akimoto & Brinton (1999: 24, 31), are provided in (195) and (196).

(195)   *Giemað þæt ge eowre* **ryhtwisnesse** *ne* **don** *beforan monnum*

'Take heed that ye [sic] do not do your righteousness before men.'
(translation of *Cura Pastoralis* 45 335.21–2)

(196)   *ac se lig…for-swælde þa ðe þa* **ceaste macedon**

'But the flame… burned up those who made the quarrel.' (*Ælfric* 221)

Example (195) illustrates the use of light verb construction *ryhtwisnesse don* 'to do righteousness', while example (196) illustrates the use of *ceast macian* 'to make a quarrel'. Some other examples are *þancas don* 'to give thanks', *answare sellan* 'to give an answer', *reste habban* 'to have a rest', and *ware niman* 'to take care' (Akimoto & Brinton 1999: 21). Akimoto & Brinton (1999) have suggested that many of the Old English collocations can be reasonably linked to Old English simplex verbs with the same meaning. This raises the question why an alternation between simplex verbs and collocation then exists, as in many linguistic situations difference in form equals difference in meaning (Langacker 1987). Following Akimoto & Brinton (1999: 50), light verb constructions bring about an extra dimension, since they tend to realise an intransitivising effect. For instance, in *cigan* 'to invoke X' and *cignesse don* 'to give an invocation' (Akimoto & Brinton 1999: 50), simplex verb *cigan* needs a direct object, whereas light verb construction *cignesse don* is used intransitively. Additionally, light verb constructions have greater "flexibility of verbal modification" (Brinton & Arnovick 2011: 2) than simplex verbs, as was discussed in Section 7.2.2., and the co-occurrence of light verb constructions and simplex verbs has also been linked to differences in stress placement (Brinton 1996: 185).

Note that the development of light verbs was a native development; however, it may have been reinforced by a similar construction in Latin, namely composite predicates with *agere* and *facere*, which both correspond to English *do* and which may have been calqued in translations (Akimoto & Brinton 1999: 54).

Although light verb constructions already existed in the Old English period, it was not until the end of the Middle English period that the usage of light verbs began to increase drastically (e.g. Görlach 2003: 97; Matsumoto 1999: 61; Traugott 1999: 259; Ronan 2014)[88]. This is illustrated by the fact that "CPs [composite predicates] seem to be the preferred expression" compared to simplex verbs, and that "there is sometimes more than one in the same sentence" (Matsumoto 1999: 61). According to Ronan (2014: 30), "the constructions are about two times more frequent in the Middle English data than in the Old English data". What is more, light verb constructions in Middle English can be used as a "poetic, stylistic device" (Matsumoto 1999: 61). Two examples of light verb constructions written by Chaucer (Matsumoto 1999: 62, 63) are presented in (197) and (198).

(197)   *For understoond now, ye asken conseil to **do wikkednesse***

     'For understand now (that) you are asking advice to commit sinful acts (lit. do wickedness).' (*Tale of Melibee*, c1390, HC)

(198)   ***Tak reward** of thy value, that thou ne be to foul to thyself*

     'Respect (lit. Take reward of) your dignity, so that you are not too wicked to yourself.' (*The Parson's Tale*, c1390, HC)

In example (197), the English-origin light verb construction *don wikkednesse* is used to express 'to do wicked action'. Example (198), next, shows the use of light verb *taken* and French-origin deverbal noun *reward* in the construction *taken reward of* 'to take account of, pay attention to, consider'. It should be noted that in this stage of the language, light verb constructions are "partially but not highly idiomaticized" (Matsumoto 1999: 92). That is,

> complex verbs seem to have had a greater degree of variability in the constituent verbs, prepositions, and nouns than in present-day English, greater interchangeability with simplex verbs, and a fuller range of nominal modifiers (Brinton & Akimoto 1999: 16–17)

---

[88] Other multi-word idioms, such as phrasal verbs and prepositional verbs, were on the rise in Middle English as well (Fischer, De Smet & Wurff 2017: 142). However, this thesis focuses solely on light verb constructions.

However, this variability of light verb constructions compared to Present-day English already started to decrease in Middle English — compared to Old English — with many constructions becoming lexicalised (Brinton & Akimoto 1999: 17).

The frequency of light verb constructions further increased in Early Modern English, and this rise continued until Late Modern English (Claridge 2000: 96–101). It is, therefore, not surprising that light verbs have worked their way into Present-day English. In fact, like *do*-support, they have become a key feature of Present-day English, two examples being *to give rise to* and *to make use of* (Akimoto & Brinton 1999: 21). The prevalence of light verbs in English is in contrast with German, which has some light verb constructions (e.g. *eine Übersetzung machen* 'to do a translation'), yet their number is much more restricted than in English (Berg 2014: 500–503). English *to do a waltz*, for example, can in German only be translated by means of a lexical verb, namely *Walzer tanzen* (lit. 'to dance waltz'). Similar to German, the Dutch language contains some light verb constructions, such as *uitsluitsel geven* ('to give a decisive answer') and *een opmerking maken* ('to make a remark') (Everaert 2008: 97; also see Verhagen 2009); however, to our knowledge the literature never provides more than a couple of instances[89]. Thus, the prevalence of light verbs in English distinguishes the language from the continental West-Germanic languages.

## 7.2.4. Light verb constructions in French

Like English, French contains a large number of light verb constructions. Examples are *avoir peur de* ('to fear') (Samvelian, Danlos & Sagot 2011: 8), *voir le jour* ('to emerge'), *prendre forme* ('to take shape') (Fuchs & Garnier 2021: 3), and *porter atteinte* ('to undermine') (Valli 2007: 46). Light verb constructions had already been part of French in earlier phases of the language as well, for instance in Middle French. Valli (2007), for example, analyses verbo-nominal expressions with nouns *garde* ('guard') and *congié* ('vacation') in 14th- and 15th- century French. Since the presence of light verb constructions is a point of compatibility between the Middle English and Middle French languages (e.g. Sankoff 1998), the use of French-origin verbs in light verb constructions in Late Middle English cannot be hindered by language-specific constraints.

## 7.2.5. Hypotheses

### 7.2.5.1. Language contact hypothesis

As discussed in Section 7.1., a set of periphrastic constructions quickly surged in usage frequency at the end of the Middle English period, and this peak coincided with the

---

[89] This is the case for Standard Dutch. Given the common association of light verb constructions with colloquial language (cf. Section 6.2.1.), we do not reject the possibility that informal or colloquial Dutch contains a higher number of light verb constructions. Moreover, in many regions in the Low Countries, Dutch is influenced by French, which has a high productivity of light verbs (cf. Section 7.2.4.).

borrowing peak of French items at the end of the 14th century (Mugglestone 2006; Baugh & Cable 2013). We argue that accommodation biases may have reinforced the drift of English towards non-finiteness. First, loans tend to enter their recipient language as lower-frequency items and, therefore, tend to be used in innovative variants, such as light verb constructions. Also, language users tend to use loan words in periphrastic structures (cf. Section 5.3.4.), and nouns are more borrowable than verbs (e.g. Muysken 1981; Poplack, Sankoff & Miller 1988; Matras 2007, 2009; Winford 2010; Seifart 2017; Hock & Joseph 2019). Light verbs, then, can serve as a tool to use loans periphrastically as well as nominally. In fact, light verb constructions are cross-linguistically known to simplify the process of loan verb accommodation (Wohlgemuth 2009: 102). Therefore, we hypothesise that French-origin deverbal nouns were overrepresented with light verb constructions compared to English-origin nouns (*language contact hypothesis*). This way, light verb constructions containing French-origin deverbal nouns may have served two functions: (i) bypass loan verb accommodation, and (ii) create new V + N combinations with various newly borrowed nouns.

Although it is unlikely that the rise of light verbs has been caused by the enormous influx of French in the second half of the 14th century — light verbs strictly speaking predated French influence (cf. Akimoto & Brinton 1999) — it may at least have been accelerated by French influx. More concretely, the large number of French words (i) may have boosted the grammaticalisation of new periphrastic strategies, such as light verbs, and (ii) may then have further promoted the reliance on already existing periphrastic strategies, light verbs included.

Our hypothesis that light verbs will be more common with French-origin nouns than with English-origin nouns is in line with the non-finite bias found in Chapter 3. While modals and auxiliaries are not light verbs in a strict sense, composite verb forms share with light verb constructions the characteristic that the syntactic and lexical functions of the verb phrase are realised separately. While the finite element (i.e. the modal or auxiliary) realises temporal and modal grounding, the non-finite element (i.e. the infinitive or participle) specifies the process type (cf. Matras 2009: 182; Myers-Scotton & Jake 2014).

Our hypothesis builds on a study by Ronan (2012; also see 2014: 23), who has conducted corpus research on the use of French and Latin deverbal nouns in light verb constructions. In a sample of Chaucer's *Canterbury Tales*, she has found that the number of foreign-derived deverbal nouns in light verb constructions is considerably higher than in Old English, and she has linked this to the higher levels of language contact (with French and Latin) in Middle English compared to Old English. She has also observed that Chaucer used light verb constructions to incorporate Romance

loan verbs in English. Although in some cases the tendency of Romance verbs towards light verb constructions seems to be due to rhyming purposes, in other cases rhyming purposes cannot account for the high proportion of French and Latin deverbal nouns in light verb constructions (Ronan 2014: 25). Additionally, she has found differences among the characters: whereas the Knight, who is educated and cultured, uses the highest proportion of foreign-derived deverbal nouns, the Miller uses the lowest proportion of foreign-derived deverbal nouns (Ronan 2012: 155). Last, she has also observed that the light verb which is most common with foreign-derived nouns in Middle English is *do* (Ronan 2012: 156). Different from Ronan's (2012) research, however, the dataset used in this chapter does not only contain Chaucerian data, but also texts by other authors. This way, we hope to show that an overrepresentation of French-origin nouns in light verb constructions is not merely a stylistic feature of Chaucer's work. Additionally, the number of different light verbs in our analysis is higher than in Ronan's (2012).

In addition, the overview of light verb constructions provided by Matsumoto (1999) and Tanabe (1999) offers another indication that light verbs in Late Middle English were frequently used with deverbal nouns of French origin. For *make*, for instance, Tanabe (1999: 106–107) lists eight possible light verb constructions: *make acowntte* ('to account'), *make affray* ('to disturb'), *make an end* ('to finish'), *make assaught* ('to assault'), *make chere* ('to have fun'), *make labore* ('to make an effort'), *make puryauns* ('to prepare'), and *make re-torne* ('to return'). From this list, the only English-origin deverbal noun is *end* in *make an end*, whereas the other seven constructions are of French origin. That the lexical components are predominantly of French origin is telling, since the English language — no matter how large the influx from French — has never predominantly contained French-origin vocabulary items. Some of the combinations mentioned by Tanabe (1999: 106–107) were directly modelled on French light verb constructions, as also seen by Iglesias-Rábade (2001; cf. infra). Examples are *taken avauntage* ('to take advantage') modelled on French *prendre avantage*, *taken beginning* ('to get started) on *prendre comencement*, *taken end* ('to end') on *prendre fin*, and *taken heart* ('to gain courage') on *prendre corage* (Iglesias-Rábade 2001: 155–156). Even where combinations were not directly modelled on French light verb constructions, their incipient success in Middle English could be linked to their ability to accommodate French-origin lexical material with verb-like functions. For *make*, for example, Tanabe (1999: 108) has suggested that "[i]t constitutes good evidence that MAKE + N + (P) tends to be used with derivative nouns of French descent". This may also be the case for light verbs other than *make*.

The impact of French on the use of light verbs in Middle English has previously also been alluded to by Prins (1952), Hiltunen (1983 in Ronan 2014: 16), and Iglesias-Rábade (2000: 96, 2001: 156). Iglesias-Rábade (2001: 156) has found that

most Old French light verb constructions "occurred in lME [Late Middle English] following the French pattern", so that "OFr [Old French] expressions such as *faire cause commune, faire mention de* were reproduced in English as 'to make common cause', 'to make mention of'". Additionally, Iglesias-Rábade (2000: 96) points to "a tendency to translate the OF 'light' verbs (*faire, prendre, porter*) into English, whereas nouns, adjectives and adverbs maintained their French form" adapted to English spelling. As such, he claims that "Middle English underwent an extensive process of romanization of its phrasal system" (Iglesias-Rábade 2001: 156), and he concludes that

> French contributed decisively to the use of this type of structure consisting of a light verb translated from French + a deverbal element which bears the action and the lexical meaning and which usually kept the French form and content. (Iglesias-Rábade 2001: 156)

However, this thesis goes one step further by claiming that not only the influx of specific French light verb constructions, but the lexical influx of French verbs in general can be held accountable for the dramatic rise of light verbs in Late Middle English.

A last argument supporting our hypothesis is that the use of light verbs in light verb constructions resembles the use of *do* in *do*-support, both being the inflected verb carrying the grammatical information. Both constructions are typical features of children's speech (Tieken-Boon van Ostade 1990), implying that they are not complex to acquire. Indeed, L2-acquisition research has shown that "light verbs such as copulae, modals, and auxiliaries are used in finite constructions earlier than lexical verbs" (Verhagen 2009: 203), which means that light verb constructions — like *do*-support — can be used as facilitative strategies for lexical verbs associated with an elevated processing cost, such as loan verbs (cf. Section 5.3.4.). Moreover, light verbs have been linked to learner interlanguages and contact situations, meaning that they tend to be less complex to acquire than simplex verbs (Danchev 2011: 30).

### 7.2.5.2. Modification hypothesis

As explained in Section 7.2.1., the primary function of light verb constructions is to modify the deverbal noun (Akimoto & Brinton 1999: 51–52; Bonial 2014). An example of adjectival modification is provided in (199), where deverbal nominal *knowing* in the light verb construction *have knowing* ('to know, have knowledge of') is accompanied by the quantitative adjective *gret* ('large, big') used in the comparative, hence *gretter*.

(199)  *thou shalt* **have** *the* <u>*gretter*</u> **knowing** *of thyn oune instrument.*

> 'You shall know (lit. have greater knowledge of) your own instrument better.' (*Treatise on the Astrolabe*, c1391, HC)

In this thesis, we hypothesise that — to the extent that light verb constructions are exploited to accommodate loan verbs (cf. Wohlgemuth 2009) — their use will tend to be less motivated by their other possible functions, such as the modification of nouns. Internal modification is not possible in, for instance, French-origin *to do execution*: attestations such as \**do **much** execution*, where the light verb is used with a quantifying adjective, and \**do **good** execution*, where the light verb is used with a qualifying adjective, are unattested. More concretely, we hypothesise that additional modification is more common in light verb constructions with English-origin deverbal nouns than in light verb constructions with French-origin deverbal nouns (*modification hypothesis*). That is because English-origin nouns have been inherited and have, therefore, been part of the English language for longer than French-origin nouns. As a result, they may also have grammaticalised more than French-origin nouns. The focus of this study will be exclusively on the adjectival phrase, which is by far the most common type of modification in light verb constructions (Moralejo-Garate 2001: §1.4.).

The language contact and modification hypotheses will now be tested by means of a corpus study. The data and methodology will be explained in Section 7.3.

## 7.3. Data and methodology

### 7.3.1. Data extraction

For this case study we extracted the data from the third sub-period of the *Helsinki Corpus of English texts* (Rissanen et al. 1991). This sub-period, called "M3", spans all texts between 1350 and 1420, and contains 140,380 words of prose, divided between 27 text samples from various genres, authors, and dialects. Poetry was not included in the dataset because metre and rhythm may distort the use of finite and non-finite (and thus periphrastic) forms (Kemenade 1987; Fischer, De Smet & Wurff 2017). Since we have also used this dataset to trace accommodation biases of French loan verbs in Middle English (cf. Section 3.3.1.1.2.), this section does not expand further on the dataset nor the data extraction procedure. An overview of the 27 text samples with their corresponding authors and word counts can also be consulted in Chapter 3. In contrast to the dataset in Section 3.3.1.1.2., however, we annotated the entire dataset instead of a 25% sample.

### 7.3.2. Data annotation

By means of close-text reading, the attestations in the sample were manually lemmatised and marked for the presence of light verb constructions (more information on which will be provided in Section 7.3.2.1.). Cases of light verb constructions were additionally annotated for the source language (English, French, and Romance) of the deverbal noun. Romance-origin nouns are included in the

category of French in the analysis. Since Latin loan words fall beyond the scope of this thesis, deverbal nouns of Latin origin, as in *take* **circumcisioun**, were not included in the sample. For a detailed description of decisions concerning source language, we refer back to Chapter 3. Additionally, we checked whether the deverbal noun of the light verb construction was modified. The construction was considered modified if it was accompanied by a quantifying adjective (e.g. *full*, *many*, *much* including spelling variations) or a qualifying adjective (e.g. *good*, *rightful*, *spiritual* including spelling variations). This means that the focus was exclusively on adjectival modification of the noun (cf. Bonial's 2014 definition): therefore, negative adverbs (e.g. *no*, *not*, *no way*, *never*) and demonstrative pronouns (e.g. *such*) were not included in the annotations. Under the scope of this thesis, relativization of the noun object (cf. Bonial's 2014 definition) was not looked at in the analysis either.

In this corpus study we analysed light verb constructions formed with the following eight light verbs[90]: *beren* 'bear'[91], *don* 'do', *given* 'give', *haven* 'have', *maken* 'make', *putten* 'put', *setten* 'set', and *taken* 'take'. The selection of these light verbs is based on the ones analysed, observed, or mentioned by Matsumoto (1999) and Tanabe (1999). The dataset included some cases, as in example (200), where two nouns (i.e. *scorne and despite*) are attested with one and the same light verb (i.e. *hath*).

(200)    *and now euery man* **hath** *of me* **scorne and despite**

'And now every man mocks and humiliates me.' (*The Brut or The Chronicles of England, c1400, HC*)

*Have scorn* is a valid light verb construction. Since *have despite* can be a light verb construction as well, and it can be dependent on *hath*, the noun is assumed to occur with an elided form of *have*.

We excluded attestations of potential light verbs being used lexically, as in examples (201) and (202), where *do* selects for additional arguments.

(201)    *to* **do** *hem þe affecioun of worschippe*

'to do him the affection of worship' (*John of Trevisa's Polychronicon, a1387, HC*)

(202)    *That wyn ne* **dooth** *to folk no swich offence.*

'That wine does to folk no such offence.' (*The Summoner's Tale, c1390, HC*)

---

[90] We also looked for light verb constructions formed with *comen* 'come', *getten* 'get', *kepen* 'keep', and *layen* 'lay', but our dataset did not contain any attestations for these potential light verbs.
[91] For reasons of clarity and because of the high degree of spelling variation in Middle English, we use the Present-day English variant of the light verb wherever possible.

In example (201), for instance, *hem* is the indirect object to *do. Dooth* ('does') in example (202) has *to folk* as its recipient argument.

### 7.3.2.1. Categorisation of light verb constructions

Categorising constructions as light verb constructions is a particularly challenging endeavour: not only can a multitude of verbs adopt the function of light verb (e.g. *do, have, make*), light verb constructions also come in different types (e.g. with or without a preposition following the noun; with or without an indefinite article). Moreover, the delimitation of light verbs and light verb constructions is challenging, since the literature contains many overlapping notions and definitions. Also, we can only look at light verbs from a Present-day English perspective, not knowing to what extent Middle English constructions were perceived as idiomatic by language users then. In Ronan's (2012: 147) words, "the degree to which collocations were fixed constructions in the earlier language is difficult to determine". For instance, Brinton & Akimoto (1999: 16) have pointed out how complex verbs "had a greater degree of variability in the constituent verbs, prepositions, and nouns than in present-day English". Particularly in the case of *have* it is not always clear whether a construction (e.g. *have mercy/compassion/shame*[92]) is a light verb construction, or an instance of *have* with a direct object. We, therefore, created a set of tests with which the attestations had to comply in order to maximise objectivity and to decrease arbitrariness in the categorisation procedure (cf. Quirk & Mulholland 1964: 68–71 for complex prepositions). The tests gauge the semantic as well as the (morpho)syntactic features of the constructions. Whereas some tests have to be complied with at any cost, most tests are not fully diagnostic and rather serve as a tool to draw the line in case of doubt. This will be explicitly mentioned for each of the tests presented below.

**Test 1.** *Is the construction idiomatic?*

In order to qualify as a light verb construction, the construction first of all has to be idiomatic. This aligns with the definition of light verb constructions as provided in Section 7.2.1. A way to test for idiomaticity is to consult the meaning of the deverbal noun in the *Middle English Dictionary* (Lewis 1952–2001), and to verify whether the deverbal noun is explicitly mentioned to occur with the light verb in question. In the case of idiomatic *take heed*, for example, noun *hed* is said to occur with *taken*, and in that case it means 'to pay attention to (sth.), mark, consider'.

**Test 2.** *Is the construction noun-based, and if so, is the noun deverbal?*

Consistent with the definition of light verb constructions (e.g. Algeo 2006), the construction should be noun-based, as in *do* **execution**. In addition, the noun should

---

[92] These attestations were not classified as light verb constructions.

be deverbal as opposed to non-deverbal, meaning that it should be derived from a verb or verb phrase. This is the case for *execution*, which is derived from *to execute*. It also eliminates noun-based expressions containing non-deverbal nouns, such as *do* **felony** and *have* **hunger**. Notice that deverbal nouns can also include nominal gerunds, as in *do* **governing** and *have* **witnessing**. This test by definition excludes all multi-word verbal expressions based on other parts of speech than nouns, such as *make* **open**, which is based on an adjective, and phrasal verbs such as *look* **back** or *break* **in**, which are based on an adverb and a preposition respectively.

**Test 3.** *Is the deverbal noun unquantifiable?*

Despite their formal resemblance, light verb constructions have to be distinguished from verbs used with a direct object. Since light verb constructions have grammaticalised and do not literally refer to the deverbal noun as a concrete instance of that noun, it follows that the noun should be unquantifiable. Therefore, in most true instances of light verb constructions, the deverbal noun cannot be pluralised (e.g. \**do* **executions**), nor can it be counted (e.g. \**do* *three* **executions**). That most light verb constructions do not take plural nouns has been shown by Moralejo-Garate (2001) in her corpus study on idiomatisation in Middle English. In cases where the noun can be pluralised and counted, it is likely used as a direct object with the verb, with the lexical use of the verb, as in *do (three)* **miracles**. For example, on the basis of this test *do* **sin** is considered an instance of a light verb, whereas *do* **sins** is considered a verb with a direct object. However, there exist exceptions to this test, such as light verb constructions *do* **reproves**, *do* **thankinges**, and *make* **amends**, where the nouns are always used in the plural.

Another way in which deverbal nouns reveal that they are unquantifiable is the absence of definite articles. In the case of *do execution*, for instance, \**do* **the** *execution* is not grammatically possible since the language user does not refer to the execution of a specific law or person (depending on the context) for a specific purpose at a specific time of the day; instead, the focus is on the general act of executing a law or a person. Therefore, *make* **the** *declaration* is not an instance of a light verb construction, since it involves a specific declaration, and the *declaration* is much more tangible than, for example, *knowledge* or *cognisance*. This is also illustrated in example (203), where *habbeth* **the** *gouernaille* is not a light verb construction, since the author describes a specific instance of *gouernaille*, namely that *of the Citee*.

(203)　*hem that* **habbeth the gouernaille** *of the Citee vnder oure lige lord the kyng by vertue of the chartre of oure franchise*

　　　　'he who governs over the city in agreement with our Lord the King by virtue of the charter of our franchise' (*Proclamations of Nicholas Brembre*, 1384, HC)

*Have* in example (203) is used transitively. An exception to this test are light verb constructions with open slots, such as *one's* in take **one's** *rest*, which has the same meaning as 'to rest', and is, therefore, a light verb construction. The decision to include constructions with open slots is in line with *call to* **one's** *mind* in Los et al. (2012: 136), which "seems to demonstrate that some degree of grammaticalization has taken place". That said, light verb constructions also regularly appear with indefinite articles (e.g. Moralejo-Garate 2001).

Whereas this test tackles the syntactic aspect of abstract deverbal nouns, the next test investigates it from a semantic point of view.

**Test 4.** *Does the deverbal noun refer to a specific, concrete instance?*

Los et al. (2012: 135–136) have found for adjectives in resultative constructions (e.g. *rinse* **clean**, *make* **happy**) that they tend to develop less concrete meanings over time, and that "the development of new meanings may have led to the underspecification of the semantics of these adjectives, and, in turn, to their grammaticalization" (Los et al. 2012: 136). Light verb constructions have a similar structure to resultative constructions, with the difference that they consist of a noun instead of an adjective. Also, light verb constructions are typically grammaticalised constructions and, therefore, the reasoning for adjectives in resultatives may apply to deverbal nouns as well. Concretely, this means that the semantics of the deverbal noun may have become underspecified over time. Underspecification of meaning has also been discussed by Koenig (1999) in his research on French subject clitic *on*, which can be used to mean *we*, *people*, or *someone*. In what he calls the "ultra-indefinite" use (Koenig 1999: 236), where *on* is used in the meaning of *someone*, the "subject [of *on*] does not satisfy an agentive semantic role" (Koenig 1999: 238). Consequently, *on* in this use cannot be referred back to, and is thus "discursively inert" (Koenig 1999: 241). This impossibility to refer back to *on* in the ultra-indefinite use is illustrated in example (204), since *on* (i.e. the person who killed the president) cannot be resumed and specified as *he* (or any other personal pronoun).

(204) *\***On**      *a*      *tué*   *la*      *présidente.*

     INDEF      have   kill    the      president

     *Il*              *était*   *du*              *Berry,  paraît-il.*

     **he**            be      from.the      Berry,  seem-it

     'Someone killed the president. He comes from the Berry, it seems.'

     or 'The president was killed. He comes from Berry, it seems.' (Koenig 1999: 241, emphasis added)

Based on the research by Los et al. (2012: 135–136) for adjectives in resultative constructions and by Koenig (1999: 241) for *on*, attestations from our dataset are only considered instances of light verb constructions if the deverbal noun does not refer to a specific, concrete instance. Allerton (2002: 242), too, has found that most nouns in stretched verb constructions are abstract. For instance, *take* **a sip** would be classified as a light verb construction, since it is "more abstract than the lexical meaning of the full verb *take* in *take* **a slice of toast**" (Elenbaas 2013: 48–49, emphasis added). Therefore, *take* **a slice of toast** is not considered a light verb construction. The same logic applies to, for example, Middle English *make* **a help**: *a help* is a concrete instance of *help* and the construction is thus not an instance of a light verb construction. This is in contrast with, for example, *make* **knowledge**, where *knowledge* is not a concrete or specific instance of *knowledge*, and which thus is a light verb construction. Such verb-noun constructions as *make* **knowledge**, which contain cognitive nouns, are all considered light verb constructions due to the abstractness of the cognitive noun. Other examples are *have* **understanding** ('to understand'), *have* **knowing** ('to know'), and *take* **cognisance** *of* ('to take notice of').

**Test 5.** *Can the construction be substituted by a simplex verb?*

Based on Dixon (1992: 340), Claridge (2000) has suggested that "two sentences, one with the simplex, the other with the verbo-nominal combination, should have essentially the same meaning[93]" for the verbo-nominal combination to be a light verb construction. Indeed, Akimoto & Brinton (1999: 50) have found "a relation of synonymity" between almost all Old English light verb constructions and their corresponding simplex verbs. Therefore, a verbo-nominal combination is considered a light verb construction if it has a corresponding simplex verb, either in English (for native English deverbal nouns), or in their language of origin (for loan deverbal nouns). An example is *do execution*, which corresponds to the simplex verb *execute*, and *do gilour*, which does not have a corresponding simplex verb in Middle English, but which corresponds to French-origin verb *guilëor* ('to deceive, beguile'). That some French-origin deverbal nouns do not have corresponding simplex verbs in Middle English ties in with the idea that nouns are more borrowable than verbs (e.g. Whitney 1881; Muysken 1981; Matras 2007), and that some French nouns entered the English language before their verbal counterparts (Ronan 2012: 152).

The light verb construction and the simplex verb having the same meaning also implies that, in both, the verb's participants map onto the same argument roles. That is, whereas one could argue that *bear crown* has a corresponding simplex verb (i.e. *corounen*) in Middle English, in *bear crown* it is the subject who bears the crown,

---

[93] Light verb constructions and simplex verbs were said to have the same meaning if one of the meanings provided in the *Middle English Dictionary* (Lewis 1952–2001) corresponded.

whereas in *corounen* it is most likely someone else, since the subject does the act of crowning.

**Test 6.** *Can the inflected verb be used in various contexts and constructions?*

According to Claridge (2000: 71), "the verbal part [of a verbo-nominal combination] is usually taken from the rather small class of very common, multi-functional verbs". Therefore, if the inflected verb in the construction is multi-functional, in the sense that it is used in various contexts and constructions, it can be considered an instance of a light verb. Classic examples are *bear, come, do, get, have, keep, lay, make,* and *take.* The semantic and functional versatility of light verbs will now be illustrated for *do* in Present-day English, which can be used as a main verb, an auxiliary verb, and a substitute verb (Cambridge Dictionary 2022b). As a main verb alone, it can be used in eight different senses (Cambridge Dictionary 2022b): (i) to perform or take part in an action (example (205)), (ii) to achieve, complete or deal with something (example (206)), (iii) for work and other tasks (example (207)), (iv) to express the study of a subject (example (208)), (v) to talk about taking part in activities (example (209)), (vi) to talk about producing or creating something (example (210)), (vii) for cleaning or tidying things (example (211)), and (viii) to state that something is enough or acceptable (example (212)). Examples for the eight meanings, cited from Cambridge Dictionary (2022b), are listed below.

(205)   *What have you been **doing** today, anything interesting?*

(206)   *She **does** <u>the crossword</u> in the newspaper every day.*

(207)   *I'm going to **do** <u>some work</u> in the garden this weekend.*

(208)   *What did you **do** at university?*

(209)   *She **did** <u>a trip</u> down the Amazon when she was in Brazil.*

(210)   *She **did** <u>a lovely painting</u> of the lake where we stayed last summer.*

(211)   *The cleaner was **doing** <u>my room</u> when I came back.*

(212)   *What size bag do you need? – A small one will **do**.*

We will not expand on the use of *do* as auxiliary verb and substitute verb. Similar semantic analyses can be conducted for the other classic examples of potential light verbs and will demonstrate the versatility of true light verbs.

***Test 7.*** *Can the inflected verb be replaced by another verb?*

Claridge (2000: 112), in her monograph on the use of multi-word verbs in Early Modern English, ascribes different preferences of multi-word verbs to different light verbs[94]. For instance, "[*m*]*ake*, *take* and *give* prefer verbo-nominal combinations, while *come*, *lay*, *go*, *send*, *put*, *bring* and *find* are mostly found in phrasal verbs". It follows that constructions with potential light verbs are light verb constructions with some verbs, but not with others. For example, while **have** *sorrow* is not considered a light verb construction, but rather an instance of *have* with direct object *sorrow*, **make** *sorrow* is considered an instance of a light verb construction, meaning 'to sorrow'. However, in other cases deverbal nouns can occur in light verb constructions with various light verbs, such as *deliberacioun*, which can occur with *have*, *take*, and *make*, and which carries the same meaning with all three verbs, namely 'to deliberate, consult, take thought' (Matsumoto 1999: 65). Also note that the use of light verbs in specific constructions may differ across regions or dialects. A classic example is the alternation between *have* and *take* in constructions such as *have/take a bath*, *have/take a look*, and *have/take a walk*. Whereas *have* is the more frequent option in British and Australian English, *take* is the more frequent option in American English (for Australian English, see Wierzbicka 1982; for British and American English, see Quirk, Greenbaum & Leech 1985). That some light verbs are interchangeable or have overlapping meanings with other verbs becomes clear from Dutch, where *doen* ('to do') can be used in the meaning of *geven* ('to give'). This is the case in, for instance, <u>*Geef mij maar een appel*</u> ('You may give me an apple'), which retains the same meaning when it is realised as <u>*Doe mij maar een appel*</u> (Los 2022 p.c.).

More than a black-on-white test to eliminate erroneous light verb constructions, this test serves as a reminder that certain deverbal nouns may also occur with other light verbs, and that while some may be light verb constructions, others are not.

### 7.3.3. Data analysis

Data analysis consisted in comparing the relative frequencies of French-origin versus English-origin deverbal nouns in light verb constructions (Section 7.4.1.), as well as the rates of French-origin and English-origin deverbal nouns with internal modification (Section 7.4.2.). The contingency tables also include the general rates of French- and English-origin words in Late Middle English, and hence serves as a baseline. For example, if the type frequency of French-origin nouns and English-origin nouns in light verb constructions with *do* was 39, the baseline represents how many of those 39 nouns would be of French and of English origin in

---

[94] She does not explicitly call them light verbs, but rather "common" (Claridge 2000: 111) or "frequent" (Claridge 2000: 113) verbs.

a normal distribution. The baseline rates for vocabulary of French origin were somewhat artificially set at 23.3% for French-origin nouns, based on the loan rates provided by Dalton-Puffer (1996: 12). 23.3% is the average of the percentages for 1350–1375 (30%), 1375–1400 (20%), and 1400–1425 (20%). Accordingly, the rates for vocabulary of English origin in the contingency tables were set at 76.7%. Although it is true that the rates provided by Dalton-Puffer (1996) do not specifically apply to nouns, we have not conducted any other case studies on loan nouns and could, therefore, not use baseline rates from our own work.

To check for significance, we additionally calculated p-values by means of the Fisher's exact test, where the significance threshold was set at $p < 0.05$. The reason for using the Fisher's exact test instead of the Chi-square test is that the datasets used in this chapter are small (cf. Levshina 2015: 214). The Fisher's exact test in this case checks whether there exists a significant effect for (i) French-origin versus English-origin deverbal nouns in light verb constructions, and for (ii) French-origin and English-origin deverbal nouns in light verb constructions versus same-origin words in the baseline. Note that we artificially presume that the datasets of light verb constructions and of baseline data have the same sample size. This may affect the p-values.

We also visualised the data for the language contact hypothesis as a mosaic plot using the R-packages "ggmosaic" (Jeppson et al. 2021) and "ggplot2" (Wickham 2016).

## 7.4. Findings

In this section we first investigate the language contact hypothesis (Section 7.4.1.) and then the modification hypothesis (Section 7.4.2.).

### 7.4.1. Language contact hypothesis

Our dataset contains 294 attestations of light verb constructions formed with *bear*, *do*, *give*, *have*, *make*, *put*, *set*, or *take*. The distribution of light verb construction tokens with French-origin and English-origin deverbal nouns is illustrated in Figure 14, where the vertical dashed line marks the baseline percentage (23.3%) of French-origin items (of all parts of speech) in the Middle English language.

*Figure 14: Distribution of French-origin and English-origin nouns (n = 294) in light verb constructions with vertical dashed line set at 23.3%.*

The vertical division in Figure 14, which is roughly in the middle of the plot, suggests that light verb constructions are considerably overrepresented with French-origin nouns compared to English-origin nouns. The numbers for nouns of both origins are depicted in Table 25.

*Table 25: Token frequency of French-origin nouns and English-origin nouns in light verb constructions (Fisher, $p < 0.001$).*

|  | LVCs (n = 294) | Baseline (n = 294) |
|---|---|---|
| French-origin nouns | 148 (50.3%) | 69 (23.3%) |
| English-origin nouns | 146 (49.7%) | 226 (76.7%) |

Whereas only 23.3% of words at the time were of French origin, the share of French-origin noun tokens in light verb constructions in our dataset is 50.3%. This means that French-origin nouns are significantly overrepresented in Middle English light verb constructions ($p < 0.001$). Since loan words tend to enter the language as lower-frequency items, and frequency differences between French-and English-origin nouns were not taken into account in Table 25, we also represented type frequencies of light verb constructions (Table 26).

*Table 26: Type frequency of French-origin nouns and English-origin nouns*
*in light verb constructions (Fisher, p < 0.001).*

|  | LVCs (n = 122) | Baseline (n = 122) |
|---|---|---|
| French-origin nouns | 80 (65.6%) | 28 (23.3%) |
| English-origin nouns | 42 (34.4%) | 94 (76.7%) |

When looking into type frequencies, the effect of source language on the use of light verb constructions becomes even clearer, since most light verb constructions attested in the dataset consist of a French-origin deverbal noun (65.6%). This corroborates the language contact hypothesis formulated in Section 7.2.5.1.

A mosaic plot of the distribution of French-origin and English-origin nouns across different light verb constructions reveals that there exist differences in distribution across the light verbs (Figure 15[95]).



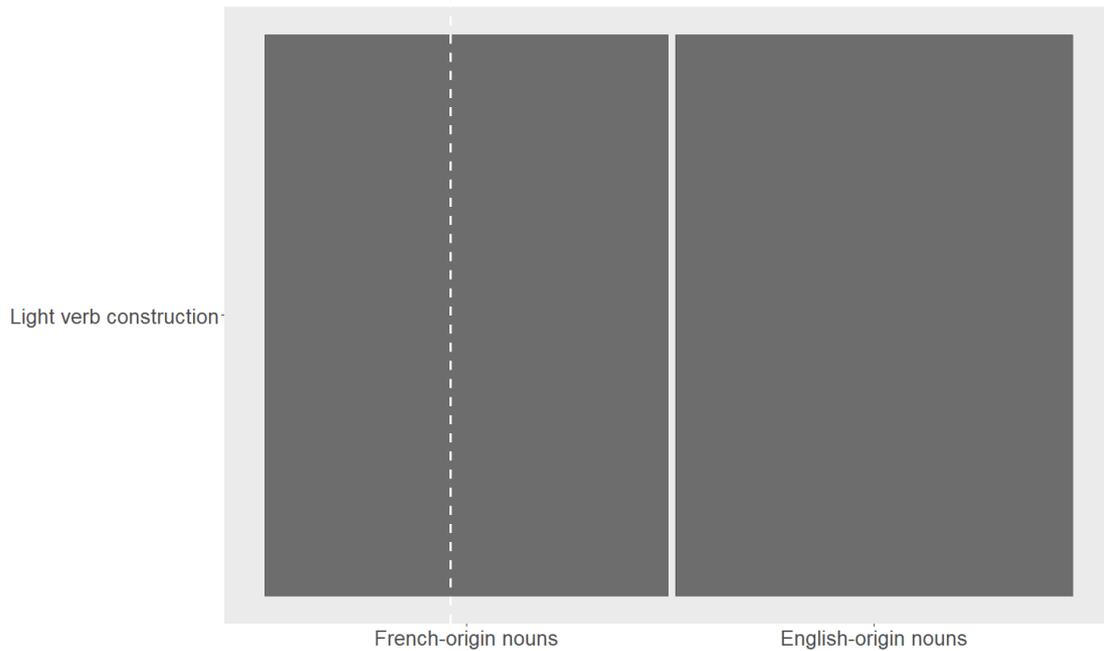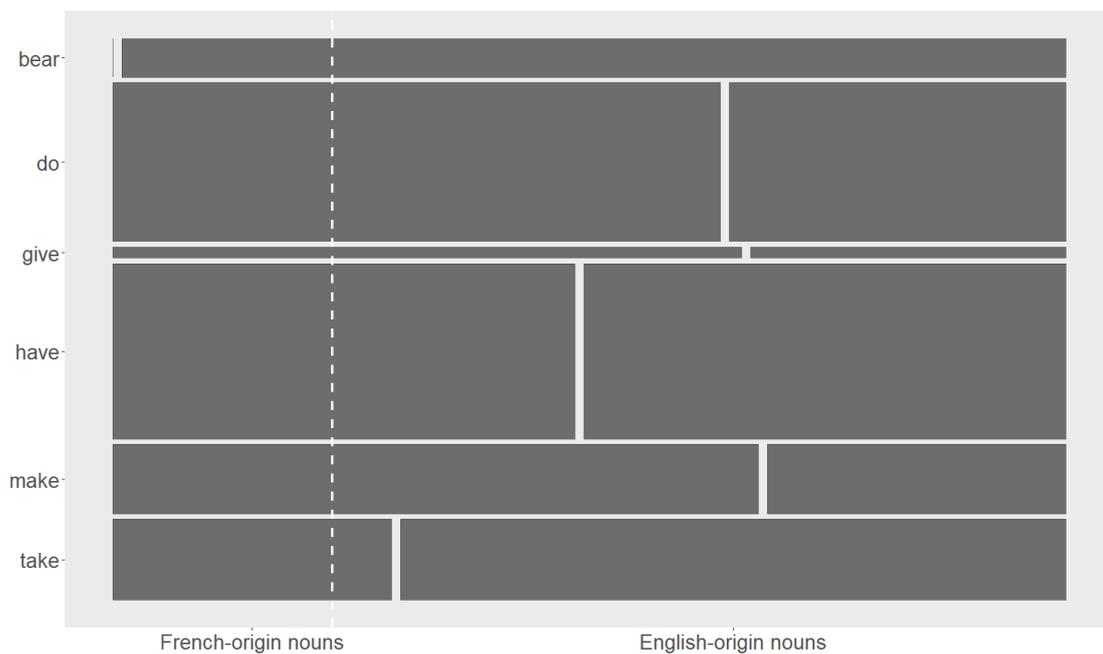*Figure 15: Distribution of French-origin and English-origin nouns (n = 294) across different light verbs with vertical dashed line set at 23.3%.*

Figure 15 reveals that French-origin nouns are overrepresented compared to English-origin nouns with all light verbs, except for *bear*. In the other cases, the

---

[95] Since *put* and *set* are attested only once, they are not included in this figure.

overrepresentation of loan nouns seems stronger for some light verbs (e.g. *do*, *give*, and *make*) than for others (e.g. *have* and *take*). However, all light verbs have different token and type frequencies, and cannot accurately be compared without looking into those in greater detail. This is why the following sections discuss the findings for each of the light verbs separately: *bear* (Section 7.4.1.1.), *do* (Section 7.4.1.2.), *give* (Section 7.4.1.3.), *have* (Section 7.4.1.4.), *make* (Section 7.4.1.5.), *put* (Section 7.4.1.6.), *set* (Section 7.4.1.7.), and *take* (Section 7.4.1.8.).

### 7.4.1.1. Bear

According to the Cambridge Dictionary (2022c), *bear* can express the following meanings: 'accept; have; support; produce; bring; change direction; say; carry; travel'. The distribution of French- and English-origin deverbal nouns with light verb *bear* is printed in Table 27.

*Table 27: Token frequency of French-origin nouns and English-origin nouns in light verb constructions with* bear *(Fisher, p = 0.05).*

|  | LVCs (n = 21) | Baseline (n = 21) |
|---|---|---|
| French-origin nouns | 0 (0%) | 5 (23.3%) |
| English-origin nouns | 21 (100%) | 16 (76.7%) |

Table 27 reveals that the number of French loan nouns in light verb constructions is not higher than the number of native English nouns; on the contrary, all of the 21 nouns in the attestations are of English origin. This goes against the general trend seen in Table 25. However, further investigation reveals that the type frequency of light verb constructions with *bear* is only 1 (Table 28).

*Table 28: Type frequency of French-origin nouns and English-origin nouns in light verb constructions with* bear *(Fisher, p = 1.00).*

|  | LVCs (n = 1) | Baseline (n = 1) |
|---|---|---|
| French-origin nouns | 0 (0%) | 0 (23.3%) |
| English-origin nouns | 1 (100%) | 1 (76.7%) |

The only light verb construction formed with *bear* is *bear witnessing* ('to witness'), a high-frequency construction containing an English-origin noun, as attested in example (213).

(213)   *And Joon **bar witnessyng**, and seide, That Y sei3 the spirit comynge doun as a culuer fro heuene*

'And Jonah witnessed (lit. bore witnessing), and said: "I saw the spirit come down like a dove from heaven".' (*The New Testament*, a1397, HC)

Due to a lack of data, the Fisher's exact test does not reveal any significant effects.

### 7.4.1.2. Do

Cambridge Dictionary (2022b) provides the following meanings for *do* as a main verb: 'perform or take part in an action; achieve, complete or deal with something; work and other tasks; study a subject; take part in activities; produce or create; clean or make tidy; be enough or acceptable' (cf. Section 7.3.2.1.). *Do* is said to be "[the] most common and most frequent semantic type of a light verb" (Wohlgemuth 2009: 104), and it is the second most common one in our dataset. Therefore, light verb constructions with *do* are also commonly referred to as the *do*-strategy (Wohlgemuth 2009: 104). Light verb constructions attested with *do* in our dataset are *do amendment* and *do amends* ('to make amends'), *do default* ('to commit an offense'), *do defence* ('to defend oneself'), *do despite* ('to defy, disobey'), *do disease* ('to trouble, harm'), *do execution* ('to put into effect or enforce (a law), execute (a person)'), *do feute* ('to acknowledge one's feudal obligation'), *do folly* ('to act foolishly'), *do gilour* ('to deceive, beguile'), *do governing* ('to govern'), *do harm* ('to harm, damage'), *do homage* ('to acknowledge one's allegiance or faithfulness'), *do honour* ('to honour'), *do lechery* ('to commit adultery, fornication'), *do merchandise* ('to buy and sell, trade'), *do mercy* ('to forgive, pardon'), *do minucion* ('to do bloodletting'), *do multitude* ('to multiply'), *do offence* ('to wrong'), *do one's cure* ('to administer a treatment to'), *do one's might* ('to do all one can'), *do one's profit* ('to be of (some) benefit, be beneficial'), *do outrage* ('to indulge in excess'), *do penance* ('to repent of one's sins, amend one's life'), *do pleasance* ('to give pleasure, be pleasing'), *do prayer* ('to pray'), *do profit* ('to be of (some) benefit, be beneficial'), *do reproves* ('to cause disgrace'), *do reverence* ('to give honour, pay respect'), *do service* ('to labour for'), *do shame* ('to cause (oneself) to feel shame'), *do sin* ('to injure (the soul) by sinning'), *do thankinges* ('to give thanks'), *do token* ('to give a sign'), *do touching* ('to touch'), *do treason* ('to be disloyal, deceitful'), *do trespass* ('to trespass'), and *do wikkedness* ('to commit slander'). Table 29 showcases the rate of French-origin versus English-origin deverbal nouns being used in light verb constructions formed with *do*.

*Table 29: Token frequency of French-origin nouns and English-origin nouns*
*in light verb constructions with* do *(Fisher, p < 0.001).*

|  | LVCs (n = 87) | Baseline (n = 87) |
|---|---|---|
| French-origin nouns | 56 (64.4%) | 20 (23.3%) |
| English-origin nouns | 31 (35.6%) | 67 (76.7%) |

According to the table, 64.4% of the nouns in light verb constructions formed with *do* are of French origin, while only 23.3% of the vocabulary at the time was of French origin. The rate of light verb constructions with English-origin nouns is only 35.6%, which means that English-origin nouns are significantly underrepresented in light verb constructions (p < 0.001). However, the two most frequently attested light verb constructions with *do*, namely *do harm* (11 attestations) and *do sin* (11 attestations), could distort the picture. We, therefore, look at the type frequencies in Table 30.

*Table 30: Type frequency of French-origin nouns and English-origin nouns*
*in light verb constructions with* do *(Fisher, p < 0.001).*

|  | LVCs (n = 39) | Baseline (n = 39) |
|---|---|---|
| French-origin nouns | 32 (82.1%) | 9 (23.3%) |
| English-origin nouns | 7 (18%) | 30 (76.7%) |

Again, the large majority of light verb constructions with *do* in the dataset is constructed using a French-origin noun (82.1%) instead of an English-origin noun (18%). An example of a light verb construction with *do* and a French loan noun, *gylerye* ('deceiver, beguiler'), is presented in (214).

(214)   *men þat loue to* **do gylerye**

'men who love to deceive (lit. do deceiver)' (*Handlyng Synne*, 1303, HC)

This strongly significant tendency for French-origin nouns to occur in light verb constructions corroborates the language contact hypothesis (p < 0.001). In line with Ronan (2012: 156), the overrepresentation of French-origin nouns with light verb *do* is stronger than with some of the other light verbs.

### 7.4.1.3. Give
Following the Cambridge Dictionary (2022d), *give* can have the following seven meanings: 'provide; cause; do; stretch; decide; offer; produce'. For *give*, we have

encountered six light verb constructions which all occur once: *give battle* ('to engage in combat'), *give counsel* ('to give permission, assent'), *give greeting* ('to greet'), *give judgement* ('to judge'), *give penance* ('to assign a penance'), and *give reckoning* ('to render an account'). The distribution of French- and English-origin nouns is presented in Table 31.

*Table 31: Token frequency of French-origin nouns and English-origin nouns*
*in attestations of light verb constructions with* give *(Fisher, p = 0.24).*

|                      | LVCs (n = 6)  | Baseline (n = 6) |
|----------------------|---------------|------------------|
| French-origin nouns  | 4 (66.7%)     | 1 (23.3%)        |
| English-origin nouns | 2 (33.3%)     | 5 (76.7%)        |

Again, the dataset for *give* contains more attestations of light verb constructions with French-origin nouns, such as *penaunce* ('penance') in example (215), than with English-origin nouns.

(215)   *He was an esy man to* **yeve penaunce***, Ther as he wiste to have a good pitaunce.*

‘He was an easy man to give the sacrament of penance to (lit. give penance), since he was known to donate to Church.’ (*The General Prologue to the Canterbury Tales*, c1390, HC)

Although this tendency points into the expected direction, it is not significant (p = 0.24). The dataset for light verb constructions with *give* is most likely too small for the Fisher's exact test to yield any significant effects.

### 7.4.1.4. Have

*Have* as a main verb can carry the following meanings (Cambridge Dictionary 2022e): 'possess; be ill; do; eat/drink; receive/allow; make happen/cause; suffer; experience; give birth; contain'. The light verb constructions in our dataset are most frequent with light verb *have*. The following constructions were attested: *have commandment* ('to give orders, instructions'), *have conning* ('to be skilful'), *have default* ('to commit an offense'), *have desire* ('to desire'), *have despite* ('to defy, disobey, resist'), *have devotion* ('to be devoted, adore'), *have disdain* ('to be indignant, take offense'), *have doubt* ('to doubt'), *have envy* ('to envy'), *have excusacioun* ('to forgive, pardon'), *have joy* ('to be joyful, be happy'), *have knowing* ('to know, have knowledge of'), *have knowledge* ('to know, have knowledge of'), *have life* ('to be alive'), *have marvel* ('to wonder about'), *have mastery* ('to prevail, win the victory'), *have mind* ('to remember, consider'), *have mistrouing* ('to lack faith, disbelief'), *have need* ('to need'), *have possession* ('to possess'), *have repentance* ('to regret, remorse, grief'), *have rest* ('to rest'), *have scorn* ('to mock'), *have solace* ('to have entertainment, enjoy

oneself'), *have sorrow* ('to be sad, suffer grief'), *have suffisaunce* ('to be sufficient'), *have thirst* ('to be thirsty'), *have understanding* ('to understand'), *have wonder* ('to inspire curiosity'), and *have worship* ('to have honour'). Table 32 presents the number of French-origin and English-origin deverbal nouns in constructions with *have*.

*Table 32: Token frequency of French-origin nouns and English-origin nouns*
*in attestations of light verb constructions with* have *(Fisher, p < 0.001).*

|  | LVCs (n = 96) | Baseline (n = 96) |
|---|---|---|
| French-origin nouns | 47 (49%) | 22 (23.3%) |
| English-origin nouns | 49 (51%) | 74 (76.7%) |

The table shows that — compared to the baseline — French-origin nouns (49%) are overrepresented in light verb constructions compared to English-origin nouns (51%), which are underrepresented. This effect is strongly significant, with p < 0.001.

Due to the presence of some highly frequent light verb constructions, such as *have knowing* (8 attestations) and *have need*[96] (11 attestations), we also looked into type frequencies (Table 33).

*Table 33: Type frequency of French-origin nouns and English-origin nouns*
*in light verb constructions with* have *(Fisher, p < 0.001).*

|  | LVCs (n = 39) | Baseline (n = 39) |
|---|---|---|
| French-origin nouns | 32 (82.1%) | 9 (23.3%) |
| English-origin nouns | 7 (18%) | 30 (76.7%) |

Table 33 reveals that 82.1% of the light verb constructions are constructed with French-origin nouns, as *enuye* ('envy') in example (216).

---

[96] The high frequency of *have need* as one of the few light verb constructions containing an English-origin deverbal noun is noteworthy considering that *need* was an impersonal verb, a construction which was disappearing at the time.

(216)  *Furst þis seed growide clene and browte forþ good fruyt, but þe feend* **hadde enuye** *þat þis seed growide þus*

'First this seed grew excellently and bore good fruit, but the enemy was envious (lit. had envy) that this seed grew so.' (*English Wycliffite Sermons*, c1400, HC)

Looking at type frequencies instead of token frequencies, the language contact effect for light verb constructions with *have* remain strongly significant (p < 0.001).

### 7.4.1.5. *Make*

The Cambridge Dictionary (2022f) provides the following meanings for *make*: 'produce; cause; cause to be; force; perform; total; calculate; earn/get; arrive; get place; appear in news'. *Make* is the fourth most frequently used light verb in our dataset, giving rise to the following light verb constructions: *make accord* ('to reconcile'), *make amends* ('to make amends, compensate'), *make assault* ('to assault'), *make assay* ('to test'), *make asseth* ('to give satisfaction'), *make avaunt* ('to boast, declare in a boastful manner'), *make comparison of* ('to compare'), *make covenant* ('to be bound by pact'), *make defence* ('to defend'), *make division* ('to divide'), *make dwelling* ('to make a sojourn'), *make estimation* ('to estimate'), *make help* ('to help'), *make homage* ('to acknowledge one's allegiance or faithfulness'), *make lesing* ('to tell a lie'), *make memory* ('to take heed, think'), *make mention* ('to write, report'), *make merchandise* ('to trade'), *make notice* ('to note, know'), *make one's prayer* ('to pray'), *make one's sacrifice* ('to sacrifice'), *make ordinance* ('to acquire, prepare for war'), *make protestation* ('to avow, declare, acknowledge'), *make purveyance* ('to make preparations, prepare'), *make refuse* ('to refuse'), *make renunciation* ('to renounce'), *make restitution* ('to return something unjustly taken'), *make shame* ('to cause (oneself) to feel shame'), *make sin* ('to injure (the soul) by sinning'), *make sorrow* ('to lament, mourn'), and *make war* ('to make war, wage a war'). The distribution of French-origin versus English-origin deverbal nouns in light verb constructions with *make* is shown in Table 34.

*Table 34: Token frequency of French-origin nouns and English-origin nouns in light verb constructions with* make *(Fisher, p < 0.001).*

|  | LVCs (n = 38) | Baseline (n = 38) |
|---|---|---|
| French-origin nouns | 26 (68.4%) | 9 (23.3%) |
| English-origin nouns | 12 (31.6%) | 29 (76.7%) |

Table 34 shows that 68.4% of the light verb constructions attested in our dataset are of French origin. This overrepresentation compared to the baseline rates for French

(23.3%), and to light verb constructions with English-origin nouns (31.6%), is strongly significant (p < 0.001). This tendency confirms our hypothesis that French loan nouns — compared to native English nouns — would be more prevalent in light verb constructions.

Due to the relatively high frequency of *make sorrow* (6 attestations), Table 35 depicts the frequencies for the types of light verb constructions with *make*.

*Table 35: Type frequency of French-origin nouns and English-origin nouns in light verb constructions with* make *(Fisher, p = 0.000).*

|  | LVCs (n = 31) | Baseline (n = 31) |
|---|---|---|
| French-origin nouns | 24 (77.4%) | 7 (23.3%) |
| English-origin nouns | 7 (22.6%) | 24 (76.7%) |

This table, again, shows that French-origin nouns (77.4%) are significantly overrepresented in light verb constructions compared the baseline of French loans (23.3%), and compared to the rate of English-origin nouns (22.6%). An example of a French loan noun in a light verb construction, namely *refus* ('refuse'), is presented in (217).

(217)   *they had* **made refus** *of hem to-forn tyme*

   'They had rejected (lit. made refuse of) them beforehand.' (*The Appeal of Thomas Usk against John Northampton*, 1384, HC)

The language contact effect for light verb constructions with *make* is significant since the p-value (0) is lower than 0.05.

### 7.4.1.6. Put

*Put* can carry the following meanings (Cambridge Dictionary 2022g): 'move; write; express; condition; judge; sail'. For *put*, the dataset only contains one attestation of a light verb construction, namely *put remedy* ('to provide a means of relief'), as reproduced in example (218).

(218)   *to the maladye of hym God* **putteth remedye** *to yeven hym rychesses.*

   'God provided relief (lit. put remedy) to his illness to give him wealth.' (*Boethius*, c1380, HC)

*Remedy* is a French-origin deverbal noun. However, the dataset is too small to draw any conclusions for light verb *put*.

### 7.4.1.7. Set

Cambridge Dictionary (2022h) provides a number of meanings for *set*: 'position; cause a condition; establish; get ready; fix/become fixed; give work; put; arrange; move down'. The dataset contains only one light verb construction with *set*, namely *set aspie* ('to scout, spy'), which occurs only once and is attested in example (219). *Aspie* is a deverbal noun of French origin.

(219)  *Y schal sette enemytees bitwixe thee and the womman, and bitwixe thi seed and hir seed; sche schal breke thin heed, and thou schalt **sette aspies** to hir heele.*

'And I will put enmity between you and the woman, and between your seed and her seed; her seed shall break your head, and you shall set ambush to (lit. spy on) her heel.' (*The Old Testament*, a1397, HC)

Due to the low frequency of light verb constructions with *set* we cannot examine the distributional properties found for deverbal nouns used with this light verb.

### 7.4.1.8. Take

*Take* is a light verb due to its many meanings, which are cited from the Cambridge Dictionary (2022i): 'remove; move; accept; hold; go with; transport; need; measure; act; think about/of; understand; catch; write; perform well; react; cheat; travel on'. *Take* is the third most frequently attested light verb in our dataset, forming the root for the following light verb constructions: *take cure* ('to pay attention'), *take feute* ('to exact a public acknowledgment of fealty'), *take heed* ('to take notice, observe, note, see'), *take homage* ('to accept a pledge of allegiance'), *take keep* ('to take notice, take heed'), *take mercy* ('to pardon, forgive'), *take one's disport* ('to amuse oneself, have fun'), *take one's rest* ('to rest'), *take pity* ('to take pity, be merciful'), *take reward* ('to take a good look'), *take shame* ('to feel shame'), *take vengeance* ('to carry out revenge'), and *take witnessing* ('to witness'). Table 36 presents the proportion of French-origin and English-origin deverbal nouns being used in light verb constructions with *take*.

Table 36: Token frequency of French-origin nouns and English-origin nouns in light verb constructions with take (Fisher, p = 0.63).

|                       | LVCs (n = 44) | Baseline (n = 44) |
| --------------------- | ------------- | ----------------- |
| French-origin nouns   | 13 (29.6%)    | 10 (23.3%)        |
| English-origin nouns  | 31 (70.5%)    | 34 (76.7%)        |

French-origin nouns are attested in light verb constructions (29.6%) slightly more often than what would be expected from the baseline (23.3%). However, this

tendency is not significant (p = 0.63). Due to the high frequencies of *take heed* (12 attestations) and *take keep* (13 attestations), Table 37 depicts type frequencies for light verb constructions with *take*.

*Table 37: Type frequency of French-origin nouns and English-origin nouns in light verb constructions with* take *(Fisher, p = 0.68).*

|  | LVCs (n = 39) | Baseline (n = 39) |
|---|---|---|
| French-origin nouns | 5 (38.5%) | 3 (23.3%) |
| English-origin nouns | 8 (61.5%) | 10 (76.7%) |

Table 37 represents that, whereas English-origin nouns make up for 61.5% of the light verb constructions in our dataset, French-origin nouns make up for 38.5%. Although the effect for type frequencies is slightly more pronounced than for token frequencies, the overrepresentation of French-origin nouns in light verb constructions with *take* is not significant (p = 0.68). This may be due to the small sample size for light verb *take*.

## 7.4.2. Modification hypothesis

A total of 59 light verb constructions in our dataset are modified, as defined in Section 7.3.3. Table 38 shows that 21% of the light verb constructions with French-origin nouns are modified versus 19.2% of the light verb constructions with English-origin nouns. Note that percentages are calculated horizontally.

*Table 38: Distribution of additional modification in French-origin nouns and English-origin nouns in light verb constructions (Fisher, p = 0.77).*

|  | Additional modification | No additional modification |
|---|---|---|
| French-origin nouns | 31 (21%) | 117 (79%) |
| English-origin nouns | 28 (19.2%) | 118 (80.8%) |

Since the proportions for additional modification in deverbal nouns of French and English origin are similar, the distribution is not statistically different (p = 0.77). Examples (220) and (221) show how a light verb construction with a French-origin and English-origin noun respectively are used with modification.

(220) *ther was* **made** *mochel* **ordinance** *þe John More, Richard Norbury, Adam Bame, William Essex, & many also mo, & þe me Thomas Vsk*

'War was well prepared for (lit. there was made much ordinance) by John More, Richard Norbury, Adam Bame, William Essex, and many more people, me [Thomas Usk] included.' (*The Appeal of Thomas Usk against John Northampton*, 1384, HC)

(221) *ȝif a man* **hadde** *ful* **knowyng** *of þis word as Crist hadde, it were but foly and veyn to here and lerne more of þis word.*

'Were a man entirely to understand (lit. had full knowing of) this word [of God] as Christ does, it would only be foolish and vain to hear and learn more about this word.' (*English Wycliffite Sermons*, c1400, HC)

Our hypothesis that modification would be considerably less common with French-origin nouns is, therefore, not corroborated. Instead, we have found that light verb constructions with French-origin nouns — once they have become part of a light verb construction — tend to be treated like those with English-origin nouns. This means that light verb constructions which are being exploited for loan verb accommodation are not limited in their other functions, in this case additional modification.

## 7.5. Discussion

In this chapter we have presented a small-scale case study on the emergence of light verbs at the end of the Middle English period. Whereas light verb constructions had already been part of the language since Old English, they only started increasing around the same time as *do*-support (cf. Chapter 6.), namely at the peak moment of French influx in the language. Loan verbs are preferred with periphrastic strategies, such as light verb constructions. Also, light verbs allow for the nominal use of loan verbs, which is an extra asset given that nouns are more borrowable than verbs (e.g. Whitney 1881; Muysken 1981; Matras 2007). We have, therefore, hypothesised that French loans would be more common in light verb constructions than native English words. It follows that the rise of light verb constructions may have been accelerated by the presence of accommodation biases in French loan verbs entering Middle English (Section 7.2.5.). As such, loan verb accommodation could be avoided, and new V + N combinations could be created with various newly borrowed nouns. This is in line with Ronan (2012, 2014), who has found in a study on Chaucerian English that "Chaucer uses more loan predicate nouns in support verb constructions than the texts in the Old English corpus do" (Ronan 2012: 158). She has attributed this tendency to the high levels of language contact during the Middle English period.

In a second hypothesis, we have proposed that light verb constructions with loans — as compared to native English words — would be less motivated by other functions, such as internal modification of the deverbal noun. That is because the light verb construction is already being exploited for loan verb accommodation. In other words, additional modification was expected to be more common in light verb constructions with English-origin deverbal nouns than in light verb constructions with French-origin deverbal nouns. Both hypotheses have been tested for the eight light verbs under investigation (i.e. *bear, do, give, have, make, put, set,* and *take*), using the 1350–1420 data from the Helsinki Corpus. The findings in Section 7.4. have shown that the distribution of English- and French-origin nouns differ slightly among the light verbs. For some light verbs, such as *put* and *set,* the sample sizes are too small to draw any conclusions. For *bear,* all deverbal nouns are of English origin, since they are all part of the same construction type, namely *bear witnessing.* Then again, with all the other light verbs French-origin nouns are consistently and significantly overrepresented as compared to English-origin nouns (Section 7.4.1.), as hypothesised. This finding does not only apply to the token frequencies of light verb constructions, but also to the type frequencies in the general dataset; however, note that the difference in distribution between nouns of English and French origin is consistently higher for type frequencies than for token frequencies, which may be due to light verb constructions with French-origin nouns typically having lower frequencies than light verb constructions with English-origin nouns. That may be because loans typically enter their recipient language as low-frequency items. The high type frequency of light verb constructions with French-origin deverbal nouns is further noteworthy since type frequency is a key determinant of productivity (cf. De Smet 2020). Therefore, the effect of type frequency must have led to a general increase in productivity of light verb constructions, which corroborates that the influx of French loan words has had a substantial effect on the rise of light verb constructions in Late Middle English.

The high rates of French-origin nouns in light verb constructions may be linked to the added processing cost of French-origin verbs in Middle English, which are not only loan verbs, but also lower-frequency items (cf. Chapter 5.). Loans tend to be used with periphrastic constructions, such as light verb constructions. Additionally, verbs are harder to borrow than nouns, and the light verb strategy may serve as a facilitative periphrastic strategy to nominalise loan verbs and thus lower the processing cost paid by language users. The findings have not provided evidence for the modification hypothesis (Section 7.4.2.); in fact, there exists no significant difference between the rates of additional modification in French-origin and English-origin deverbal nouns. Therefore, we have concluded that French-origin nouns — once they have become part of a light verb construction — are treated equally to English-origin nouns.

In sum, this case study has demonstrated that the strong rise of light verbs in Late Middle English may have been linked to a preference to use French-origin verbs periphrastically and nominally. Similar to *do*-support, it is unlikely that French influx has been the only factor to have played a role in the rise of light verbs, since light verbs already existed in Old English and, therefore, predated French influx. Even before French started influencing English, language-internal factors had given rise to the existence of light verbs. That said, the accommodation of French loan verbs may have played its part in promoting the development (cf. Thomason 2001: 62).

When interpreting these findings, one should bear in mind that this study is based on a limited amount of data. However, despite the relatively small sample sizes, the findings are significant, and the differences between words of French and English origin consistent. Also, different from the other case studies[97] in this thesis, this corpus study has not controlled the findings for potential text and register effects. Although we have only included data retrieved from prose texts, we have not considered the possibility of register differences within prose texts. The texts in the sample have been extracted from a wide range of prose texts (more formal and more informal, religious and non-religious, fictional and non-fictional), but token frequencies are inevitably higher in some texts than in others. Register effects have also been extensively discussed in, for instance, Biber & Gray (2013) and Smitterberg & Kytö (2015). Another caveat of this study is that the delimitation of light verb constructions can be vague, and that we had to make somewhat artificial decisions as to what to include and what not. However, objectivity was maximised by creating seven tests used in the categorisation process which were diagnostic of constructions being light verb constructions. A last caveat is that many of the texts in the present analyses have been translated from French and Latin originals. In their research on the translation of English light verbs into Arabic, Hassan & Malallah (2020) have concluded that translating light verb constructions is more challenging than translating simplex verbs. Therefore, many translators do not manage to convey the original meaning of light verb constructions from the original text. The role of using translated texts will be further discussed in Section 8.2.

The findings in this chapter have provided further evidence of loan word accommodation biases affecting their recipient-language grammar in the long run. Specifically, we have proven that French has exerted *indirect* influence on Middle English by showing why the English language has seen a rise in light verb constructions. The high productivity of light verb constructions in English is in contrast with the continental West-Germanic languages, where the use of light verbs

---

[97] In Chapters 3. and 4. text was included in the regression analyses as a random effect. In Chapter 6., the sample was text-matched.

is considerably less developed (for German, see Berg 2014; for Dutch, see Everaert 2008; Verhagen 2009). The findings raise the question whether it may even be the case that light verb uses of loan nouns have paved the way for the appearance of simplex verb uses of those loans. Concretely, the past use of a light verb construction such as *give rebuke* (i.e. *gave rebuke*) would predate the past use of a simplex verb such as *rebuke* (i.e. *rebuked*). This hypothesis would be in line with borrowability hierarchies, according to which nouns are more borrowable than verbs (e.g. Whitney 1881; Muysken 1981; Matras 2007). After having been borrowed, the noun can then gradually evolve into a verb, as found by Whitney (1881).

This chapter concludes Part III of this thesis, which has investigated whether the accommodation of French loan words may have diachronically altered the English grammar. Chapters 6. and 7. have demonstrated that the constant and enormous influx of French has indeed exerted influence on the strong rise of two verbal periphrastic constructions in Late Middle English and in Early Modern English, namely *do*-support and light verbs. These *indirect* language contact effects are visible to date, since German and Dutch — the sister languages of English — have never been in contact with French as intensely as English, and their share of light verb constructions is considerably lower.

**Part IV**

**Conclusions**

# Chapter 8

# Conclusions

## 8.1. Summary and implications

This PhD thesis has investigated constraints on loan word accommodation in two different parts of speech, and indirect syntactic change induced by language contact. This was done by conducting a set of case studies applied to two contact settings: English loan words in Present-day Dutch and French loan words in Middle English. Two overarching research questions have been set out to answer:

(i)    Are English loan words in Present-day Dutch and French loan words in Middle English biased towards specific morphosyntactic contexts in their recipient language? How do such biases manifest themselves, and are they persistent through time? Additionally, what factors are responsible for the occurrence of accommodation biases? (Part II)

(ii)   Do accommodation biases have long-term impact on the grammar of their recipient language? More specifically, has the great influx of French loans in English had lasting impact on some of the long-standing trends in the history of English? For instance, has the non-finite bias in French loan verbs promoted the overall use of non-finite forms in English? (Part III)

We have tackled the research questions by comparing the distributional properties of loan words and native words across different usage categories and syntactic positions. The focus was not on individual differences between language users (i.e. entrenchment), but on the integration of loan words in the language system (i.e. conventionalisation). Concretely, we have compared the rates of loan and native words in finite versus non-finite categories, attributive versus predicative syntactic position, with or without inflectional endings, and with or without *do*-support or light verbs. To this end, we have used linguistic corpus data: a corpus of spoken data for Dutch, and prose texts from a wide range of genres for Middle English. To our knowledge, the possibility of *indirect* syntactic change through probabilistic biases in loan word accommodation has never been systematically investigated in linguistic

research before. Therefore, the findings may contribute considerably to the fields of contact linguistics, historical linguistics, and cognitive linguistics.

In Part I we have provided an elaborate state of the art focusing on borrowing as a result of language contact, with a specific focus on lexical borrowings. Loan words are accommodated to their recipient language, for which we have honed in on the four loan verb accommodation strategies identified in typological research by Wichmann & Wohlgemuth (2008) and Wohlgemuth (2009). Next, we have shown that language contact can bring about contact-induced change, and that the outcome of this (lack of) change is dependent on a number of factors. Then we have introduced the two contact situations under investigation: English influence on Dutch in the Low Countries and French influence on Middle English. Although the language pairs of both contact settings are similar, their historical contexts differ considerably. Last, we have set out the general aims and hypotheses of this thesis.

In Part II we have extensively documented the existence and nature of 'loan word accommodation biases', or statistical biases of loan words towards specific morphosyntactic contexts. That loan words come with probabilistic constraints means that language users are reluctant to use lexical borrowings in some syntactic environments. Biases have been found in two parts of speech, namely verbs (Chapter 3.) and adjectives (Chapter 4.). A pilot study by De Smet (2014) had already demonstrated that the use of English loan verbs in Dutch is probabilistically biased to certain categories compared to native Dutch verbs. We have now corroborated this effect for English loan verbs, and have found a similar tendency English loan adjectives. Additionally, we have demonstrated the presence of biases in French loan words accommodating to Late Middle English. Concretely, loan verbs in both contact settings are biased towards non-finite and uninflected categories, and loan adjectives are biased towards predicative syntactic position and uninflected categories. The syntactic biases (i.e. *non-finite* and *predicative biases*) are consistently stronger than the morphological biases (i.e. *markedness biases* in both parts of speech), which sheds light on the exact nature of accommodation biases. In addition, this part has focused on the persistence and the causes of biases. Chapter 4. has shown that for French-origin adjectives in Middle English the retention of biases in foreign lexical material is remarkably persistent: biases found in loan words in Late Middle English appear as of their first occurrences in Early Middle English. Chapter 5., next, has discussed the potential causes of accommodation biases by focusing on the processing constraints associated with the morphosyntactic integration of loan words. It has been argued that loan words boost the activation of their source language in bilingual speakers, and that this leads to greater levels of cognitive effort spent on monitoring language selection. The increased processing cost can be compensated by relying on loan word accommodation biases, which act as facilitatory periphrastic strategies. This finding

has implications for the long-standing debate on the relationship between lexical borrowing and code-switching (conform with the definitions used in Sections 5.4.1. and 5.4.2.). Accommodation biases reveal that even highly integrated material is not entirely treated like recipient-language material, since inserted items do not occur in all conceivable environments with the same ease. Moreover, accommodation biases come with morphosyntactic constraints also observed in synchronic code-switching research (Myers-Scotton & Jake 2014; also see Keller 2020: 66). Hence, biases show similarities with single-word code-switches. However, biases also resemble borrowings in that foreign material is morphosyntactically integrated in its recipient language and becomes highly conventionalised over time. Accommodation biases can thus be situated in between code-switches and borrowings, marking the long and unidirectional transition from single-word code-switch to highly conventionalised borrowing. We have, therefore, proposed that code-switching and borrowing are tightly interwoven phenomena which form part of the same cline (cf. Myers-Scotton 1993b; Treffers-Daller 2005; Matras 2009; Backus 2015). Note that, although the increased processing cost when integrating loan words in recipient-language syntax is a potential cause of accommodation biases, it may not be the only one. We have, for instance, stressed the need to look at incompatibly between language pairs, and to take context-specific and language-external factors into account.

The findings of Part II have critical implications for linguistic research. It is true that loan words in direct insertion, where recipient-language inflections are added directly on the loan stem, can be treated like native words (Wohlgemuth 2009). However, our findings imply that — even under direct insertion — the accommodation of loan words comes with constraints, since loan words cannot be integrated in all usage categories with the same ease. This shows that inflection can hinder verbal borrowing, which is a small but crucial correction to Wohlgemuth's (2009) stance that inflection does not constrain verbal borrowing, as it is so commonly added to loan words through direct insertion cross-linguistically. Therefore, distributional properties of loan words in their recipient language may contain valuable information on how attached the words still are to their source language in language users' minds. Our findings also considerably improve our insight into constraints on and strategies for loan word accommodation.

In Part III we have investigated the possibility that loan word accommodation biases have prolonged consequences on their recipient-language grammar. More concretely, we have demonstrated the indirect effects of contact with French on the history of the English grammar. Many authors have suggested that, although French has led to great lexical transfer into Middle English, syntactic transfer is highly limited or even inexistent (e.g. Fischer 2013 and defenders). The possibility that French has impacted

English beyond the lexical level has, therefore, traditionally been dismissed. However, those authors have not considered the possibility that Middle English syntax may have been affected *indirectly* by contact with French: it is conceivable that loan word accommodation biases coming with French lexical loans may have reinforced ongoing internal developments and expansions (cf. Hockett 1969: 414; Thomason 2001: 62) in (Late) Middle English and its immediate aftermath. Concretely, the dramatic rise of non-finite forms (and periphrastic forms in general) and rapid inflectional loss coincided with the enormous peak of French influx. We have, therefore, hypothesised that the non-finite bias in French loan verbs may have accelerated the rise of non-finite forms in Late Middle English (1350–1500) and Early Modern English, which immediately followed Late Middle English. This may have resulted in an overall drift towards greater reliance on non-finite verb forms, including the rise of verbal periphrastic constructions such as *do*-support (Chapter 6.) and light verb constructions (Chapter 7.). Chapter 6. has indeed revealed that French-origin verbs are significantly more common with *do*-support than English-origin verbs. Loan words were also expected to be more commonly used with light verbs as compared to native words (cf. Ronan's 2014 study). Light verbs entail a periphrastic use, as in *to take heed*, and the nominal use of verbs. Thus, by using loan words in light verb constructions they become less costly, because nouns are considered to be more borrowable than verbs (e.g. Muysken 1981; Poplack, Sankoff & Miller 1988; Matras 2007, 2009; Winford 2010; Hock & Joseph 2019). Chapter 7. has corroborated the hypothesis that French loan words are more prevalent in light verb constructions than native English words. Despite the above findings, it is unlikely that French influx (alone) has caused the rise of *do*-support in 16th-century English and of light verbs in late 14th-century English. For *do*-support, first, it has been shown that a multiplicity of factors is involved in the variation between the presence and absence of the construction. In fact, *do*-support had already started to develop due to a variety of reasons — both internal and external — such as word order changes in the Middle English period. For light verbs, second, their use strictly predated French influx, as they already sporadically appeared in Old English. What is probable, however, is that French influx has triggered the rise of both periphrastic verbal constructions, since accommodation biases interacted with and contributed to syntactic change in the English language. Therefore, we may conclude that the great influx of French loans accommodating to English has had long-term impact on changes and developments in the English language. More generally, this means that accommodation biases can impact their recipient-language grammar and that they may be utilised to explain syntactic developments in languages subject to language contact.

The findings of Part III have two major implications for linguistic research. First, the use of *French verbs* in *English phrases* shows that loan word accommodation and

syntactic change in Middle English interacted. The evidence of contact-induced change in Middle English shows that French may not have directly transferred syntactic structures into Middle English, but that the rise of *do*-support in Early Modern English and of light verb constructions in Late Middle English result from lexical transfer. Therefore, contact-induced syntactic change can be indirect instead of direct, an aspect which has often been neglected in language contact research. This considerably improves our understanding of the indirect syntactic effects of language contact. Second, English differs from Dutch and German in its high productivity of non-finite forms, such as the existence of *do*-support and the high frequency of light verb constructions. The remarkably high productivity of non-finite use and possibly other striking syntactic differences between English and continental West-Germanic languages (e.g. McWhorter 2002) may be an indirect result of intense language contact with French in the Middle Ages. In other words, contact with French may have created a chasm between English and the other West-Germanic languages, which fundamentally deepens our understanding of the history of the English language and the role of French in it.

## 8.2. Caveats

The research presented here comes with a number of caveats. Whereas limitations specific to the particular case studies have already been discussed in the previous chapters, this section provides some more general limitations.

A first set of caveats pertains to the data and methodology used in this thesis. First, the study on Middle English is diachronic, and working with historical data unavoidably comes with some limitations. For instance, not all social strata and genres are represented (Davidse & De Smet 2020). Also, analysing historical data involves analysing languages in different periods in time, spoken by people with different backgrounds and mindsets (Meillet 1921: 4).

Second, some of our findings are based on relatively small sample sizes. This is particularly true for the datasets for light verbs, the distribution of *-e(n)-* and zero-endings for loan and control verbs used in the past singular and plural, and also for the distribution of syntactic position for French-origin adjectives in Early and Late Middle English. There are options to expand the sample sizes for the Early and Late Middle English periods, for example by including poetry data from the *Parsed Corpus of Middle English Poetry*, abbreviated PCMEP (Zimmerman 2018). This corpus currently contains 51 Middle English poems and can contribute in closing the data gap in the M2 sub-period (1250–1350) of the PPCME2 (cf. Zimmerman 2018). The data for the Late Middle English and Early Modern English periods used in this thesis may additionally be complemented with data from the *Parsed Corpus of Early English*

*Correspondence*, abbreviated PCEEC (Taylor et al. 2006), which is the parsed version of the older CEEC (Nevalainen et al. 1998). The PCEEC contains published authentic correspondence on mixed topics written in an informal and formal register, but with a focus on the informal one. However, implementing PCEEC data should be done with caution due to the small overlap between PPCEME and PCEEC, as described by Taylor et al. (2006). In a new study, we have started expanding the sample used to test the non-finite bias in French loan verbs (Elter & Shaw 2022). This was done by conducting a full-text analysis of the PPCME2 and the *Parsed Linguistic Atlas of Early Middle English*, abbreviated PLAEME (Truswell et al. 2018), a diachronic corpus which contains prose and poetry data from between 1150 and 1225. For PPCME2 this has yielded 92,688 attestations, and for PLAEME 17,477. This larger dataset (n = 110,165) has revealed a strongly significant non-finite bias for French loan verbs as compared to the English baseline, and this finding applied both to the entire dataset and to each sub-period of Middle English separately (i.e. M1: 1150–1250, M2: 1250–1350, M3: 1350–1420, M4: 1420–1500) (Elter & Shaw 2022). Thus, our findings for the non-finite bias in the French-Middle English contact setting hold when the sample size is scaled up.

A third caveat relating to our methodology is the artificial way of classifying loan words of Romance origin. For instance, when loans are indicated in the MED (Lewis 1952–2001) as being of shared French and Latin origin, we have included them in the category of French loans. All loans which could possibly be linked to French were classified as French. It is practically impossible to manually distinguish all loan words of French and Latin origin (as also seen in Schendl 2012: 512), and of different varieties of French, and most dictionary makers do not endeavour to distinguish them either. Following Brinton & Arnovick (2011: 252), "[o]ften it is difficult to determine whether a word was borrowed directly from Latin or came into English via French". That is because Latin influence was partially contemporary with French influence (Onions, Friedrichsen & Burchfield 1966: viii). Therefore, whether a loan word is of French origin (or a specific variety of French) is — in the analysis — less important than that it is a loan word at all. In this way, we follow Dalton-Puffer (1996: 11) in that "a French loan-word is a word which whatever may be its etymology or ultimate origin has been immediately borrowed from the French". Although this approach is sub-optimal, accommodation biases similar to the ones for French have also been found in English loans in Present-day Dutch, which strengthens the validity of the analysis.

Fourth, one of the research topics of this thesis is the use of inflection in early stages of English. However, the data used have been extracted from diplomatic text editions, and not original manuscripts. In diplomatic text editions the aim is "to make a text accessible to a larger readership" (Keller 2020: 56), which means that editors freely

make adjustments to the original manuscript. For example, editors can "adjust anything that they consider to be a common mistake or slip of the pen in the original" (Keller 2020: 56), hence also in the use of inflections. Keller (2020: 56) has underlined the importance of using 'unpolished' manuscripts in morphosyntactic research: "a scholarly edition of a handwritten text is not intended as a basis for a precise morphosyntactic analysis". Before her, Kristensson (1997) as well has recommended using written records which are as close to their original manuscripts as possible. Therefore, the markedness biases may apply to the loan verbs and adjectives in the specific (diplomatic) text editions used in this thesis, but we should consider the possibility that inflectional usage in some of the original manuscripts may well differ. It is impossible to predict whether biases in the original manuscripts would be stronger, weaker, or similar in strength. However, we conducted a spot test for *The Parson's Tale*, in which we compared adjectival inflections of the diplomatic edition (PPCME2) with those in three manuscript editions drawn from the *Corpus of Middle English Prose and Verse* (McSparran 2000): *The Ellesmere manuscript of Chaucer's Canterbury tales* and *The Hengwrt ms of Chaucer's Canterbury tales* (ed. F.J. Furnivall), and a manuscript of *The Canterbury Tales* (ed. F.N. Robinson). The manuscripts were compared to the diplomatic text edition in nine randomly selected places. In one place, the adjectival inflections in the diplomatic editions and manuscripts differed for 31.3% (5 out of 16), yet in the other eight places the difference rate stayed within a range of 0% to 15%. Overall, we did not find any troubling systematicities for adjectives of a specific origin or in a specific syntactic position to be inflected differently in the manuscripts than in the diplomatic texts. On a more optimistic note, the diachronic findings for Middle English resemble the synchronic findings for Dutch, which have not been based on translations.

A second set of caveats pertains to the interpretation of our findings. First, many of the texts investigated for Middle English and Early Modern English are translations, which are subject to a number of *translation universals* (cf. Baker 1996; Kranich, Becher & Höder 2011; Zanettin 2013). Such translation universals are "supposedly invariant features which characterize all translated texts independently of source language and translation direction" (Zanettin 2013: 21), and examples are the simplification and explicitation of the items to be translated (Baker 1996). In this thesis, we have not taken into account the possibility of *translation effects*, or the transference of (syntactic) features from source into recipient language during a translation process (Taylor 2008: 342). This means that effects found in research may depend and vary based on the data used. According to Haeberli (2010: 301), the problem is that "distinctive features of a translation could be the result of influence of the source text rather than the sign of a fundamental property of the language examined". To illustrate, in diachronic research on the position of the adjective in Middle English, Trips (2014) has found

that the number of postponed adjectives in texts based on French or Latin manuscripts is considerably higher than in originally English texts. The feature may have been borrowed from Old French and may, therefore, be more markedly present in texts translated from or based on French texts. Other examples of translation effects have been found for Romance texts translated from Latin during the Golden Age (del Rey Quesada 2017) and for the absolute construction in Old English texts translated from Latin (Pol 2016). Our own dataset contains texts which have been written under diverse circumstances: some have been originally written in English (e.g. *A Late Middle English Treatise on Horses*), while others have been translated from French (e.g. *Mandeville's Travels* and *Tale of Melibee*) or Latin (e.g. *The Old and New Testaments*). In some cases texts are said to be 'based' on a foreign language, but the extent to which such texts mirror their original texts is unclear: are resemblances purely content-related (i.e. paraphrases) or also structural (i.e. direct translations)? To further complicate matters, there are also considerable differences between the proficiency levels of French and Latin among authors in Medieval England. For example, whereas it is common consensus that Chaucer was highly proficient in French (e.g. Lumiansky 2019: para. 4), the author of *Mandeville's Travels* is said to "often misunderstand[…] the French text" (Kroch & Taylor 2000). Moreover, Taylor (2008: 356) has shown that "different constructions are differentially affected by translation". The methodological implication is that the meta-information of written texts should be meticulously investigated, and all texts which are translations (i.e. the majority of Late Middle English texts) should be compared with their original manuscripts. Only this way translation-induced contact can be discarded as a possible source of influence on the current findings (Haeberli 2010: 301). Although we have not consulted the original manuscripts in this thesis, we have analysed various texts from different backgrounds in order to mitigate possible background-specific effects. In addition, text-specific effects, which partially include the effects of translation, have been controlled for in the analyses. We would also like to highlight that translations have many assets as well, as can be seen from del Rey Quesada (2017: 674–675): not only are translators instigators of language change, typical features of the language also tend to be exaggerated in translated texts, which means that the features become easier to study.

Second and last, the actuation problem (cf. Weinreich, Labov & Herzog 1968: 102; Section 6.5.) remains as to why French loan verbs and adjectives in Late Middle English were favoured in uninflected forms. Inflectional use in that period had already started to decline drastically (e.g. Pyles & Algeo 1982: 153–154; Rissanen 2000), so the question is why loan words were still avoided in inflected contexts.

## 8.3. Avenues for future research

The present findings open various avenues for future research.

Relating to the existence and nature of loan word accommodation biases, a possible avenue is to keep focusing on the contact settings under investigation in this thesis, following the methods developed for this purpose. As mentioned in Section 8.2., the analyses in this thesis would benefit from larger sample sizes. For instance, the corpus data for Present-day Dutch could be complemented with data from Twitter, a social networking platform where users continuously share short messages and, therefore, produce instances of authentic language. Moreover, since input is constant, Twitter data allow for an investigation of the most recent language developments. The Dutch corpus data could also be complemented with online or in-person questionnaires (e.g. Schellinck 2011), containing authentic attestations of loan words in a number of morphosyntactic contexts (cf. parameters tested in this thesis). More concretely, we recommend looking for attestations of loan verbs in uninflected finite (222) and inflected finite forms (223), and in two types of non-finite forms, such as infinitives (224) and past participles (225).

(222)  *9 uur : Opstaan , wassen , ontbijten . Ik **print** een reeks documenten voor het theatre*

'9 o'clock: get up, wash, have breakfast. I print a set of documents for the theatre.' (SoNaR)

(223)  *Hij **printte** zo traag dat je mee kon lezen.*

'He printed so slowly that you could read along.' (Twitter, 24 May 2019)

(224)  *wie komt er volgende week zorgen dat ik **kan** **printen** op kot?*

'Who is going to make sure that I can print at my dorm next week?' (Twitter, 19 September 2012)

(225)  *Ik **heb** dit voor jou **geprint** .*

'I have printed this for you.' (SoNaR)

Additionally, we recommend implementing attestations of loan adjectives in attributive (226) and predicative syntactic position (227), and — within attributive position — used with (228) and without inflection (229).

(226)  *Wacht even . Je maakt deel uit van een **select** clubje .*

'Wait a minute. You are part of a select club.' (SoNaR)

(227) *Zowel de spreker als zijn publiek <u>is</u>* **select** *.*

'Both the speaker and his audience are select.' (SoNaR)

(228) *Het zijn mensen die graag bij een* **select<u>e</u>** <u>groep</u> *van levensgenieters behoren .*

'They are people who like to belong to a select group of hedonists.' (SoNaR)

(229) *Op 15 november stelt Kylie Minogue haar nieuwe album " Body Language " voor aan een* **select** <u>publiek</u> *in Londen .*

'On 15 November Kylie Minogue will be presenting her new album "Body Language" to a select audience in London.' (SoNaR)

The data can best be extracted from written sources, such as the *Stevin Nederlandstalig Referentie Corpus*, abbreviated SoNaR (Dutch Language Institute 2015). In this experimental design, it is more natural than using spoken sources, since participants are supposed to read the sentences. Sentences should not contain any loan words other than the target word (i.e. the in-context loan word) and should be clear without additional context. Respondents should assess the target words for their acceptability, which may provide insights into the degree of entrenchment (cf. Backus 2015: 27). Of course respondents should not be informed about the aim of the study in order to avoid biasing the findings. The usage frequency of loan words should additionally be controlled for in the analysis, and each participant should be provided with a unique questionnaire in order to mitigate lemma effects or unpredictable effects relating to the context of the loan words. Given the non-finite, predicative, and markedness biases found in this thesis, it is expected that loan words will be evaluated as less acceptable in finite forms, attributive position, and with inflection. If that hypothesis is borne out, the questionnaire will have shown that English loans in authentic Dutch utterances used today are subject to morphosyntactic constraints.

As for English, future studies may additionally examine whether Latin loans were subject to accommodation biases as well. It has been noted, for instance, that 15th-century Latin loan verbs in *-ate* did not always receive *-ed*-inflection in the past participle (Gąsiorowski 1997: 158–159). This is illustrated in example (230), where *sophisticat* does not receive *-ed* as a marker of the past participle.

(230) *the bawme þat is* **sophisticat**

'the balm which is sophisticated' (*Mandeville's Travels*, c1400, PPCME2)

In archaic language, the diverging use of the past participle of Latin verbs in *-ate* continued to persist until the 19th century (Gąsiorowski 1997: 158). This may be an indication of a markedness bias in Latin-origin verbs in *-ate* entering English.

Future research should also focus on other contact settings than the ones discussed here, as to assess whether loan word accommodation biases occur only in Dutch and Middle English, or whether they reveal a cross-linguistic trend. A feasible option would be to investigate the use of accommodation biases in the English-Spanish contact setting, for which an enormous collection of data is available. However, it would be all the more interesting to investigate whether biases occur in contact settings between understudied languages, and specifically in typologically distant ones. If biases are found in such contact settings as well, we may suggest that accommodation biases occur across languages, adding to tentative evidence for Kriol loan verbs entering Jaminjung (Schultze-Berndt 2017: 265). Another promising contact setting is the English-Afrikaans setting (e.g. Coetsem 1988: 129), for example by retrieving data from the *Korpus Gesproke Afrikaans* (Spoken Afrikaans Corpus) (Van Rooy 2002). Most South African citizens are bilingual, and Afrikaans has high borrowing rates from English (Deumert 2005; Van Dulm 2009). Similar to English, Afrikaans has only limited inflectional morphology left (cf. Kotzé 2001). This is interesting, since English loan verbs and adjectives in Present-day Dutch enter a recipient-language system with richer morphology than their source language. The opposite holds true for French loans in Middle English, which come from a source language with a richer morphology than their recipient language and which can, therefore, shed some of their inflections. Since the morphology of both English and Afrikaans is similar in richness, this factor cannot interfere with the biases. Therefore, the non-finite and predicative biases may tested without the interference of inflection which is present in the English-Dutch and French-Middle English contact settings. In follow-up research on the non-finite bias, we have looked at contact between Old Norse and Middle English (Elter & Shaw 2022). Although the two languages have been in contact between ca. 787 and 1042, and contact had already subsided by the end of the Old English period, many loans were only first recorded in Middle English (e.g. Hug 1987). The findings for this contact setting have revealed that Old Norse verbs in Middle English are subject to a non-finite bias as compared to the English baseline. However, the biases in Old Norse verbs are consistently less strong than those in French verbs. This difference in strength has been linked to Old Norse and Middle English being typologically and lexically closer than French and Middle English. Furthermore, the temporal distance to the period of direct contact between Old Norse and Old English is considerably larger than that between French and Middle English (Elter & Shaw 2022). One should also bear in mind that the two language contact settings have various sociological and historical differences (e.g. Townend 2002; Ingham 2020).

Regarding the causes of accommodation biases, it may be worthwhile to assess which other factors cause the existence and persistence of biases. It is also highly

recommended to verify experimentally whether the morphosyntactic integration of loan words comes with an increased processing cost, as we have set out in this thesis. This could be tested by means of pupillometry, as suggested in Section 5.6. Both avenues have been discussed in detail in Chapter 5.

A last valuable avenue for future research relates to the long-term impact which accommodation biases can have on the grammar of their recipient language. It is plausible that the impact of French influx on the history of the English language was not limited to the developments discussed in this thesis. For instance, it may be relevant to investigate periphrastic developments, other than *do*-support and light verbs, from the time when French influx peaked. A construction which may have expanded by finiteness avoidance in verbs of French origin and which has emerged in Late Middle English is the gerund. It is a characteristic feature of Present-day English, in contrast with German and Dutch, where its use is more restricted (Hantson 2003: 78; Berg 2014: 506–508). Formally it is composed of the verbal stem plus *-ing* (including spelling variants). According to Jack (1988), Fanego (2004), and De Smet (2008), the gerund had an older nominal variant, as shown in example (231), which had already been part of the language since the Old English period. Throughout Middle English, the nominal gerund started to verbalize and evolved into a non-finite clausal (also called verbal) type, as shown in example (232).

(231)  *Understond wel that evermo, fro the* **arising** *of the sonne til it go to reste, the nadir of the sonne shal shewe the houre of the planete*

‘Understand well that forever and ever, from the rising of the sun until it goes to rest, the nadir of the sun will show the planetary hours.’ (*Treatise on the Astrolabe*, c1391, HC)

(232)  *Thanne is discipline eek … in* **suffrynge** *paciently wronges that been doon to thee*

‘Then discipline also resides in patiently suffering the wrongs done to you.’ (*The Parson's Tale*, c1390, PPCME2)

Nominal gerunds cannot only be combined with *of*-phrases (*of the sonne* in (231)) and article (*the* in (231)), but also with adjectives; verbal gerunds can be combined with common-case subjects, objects (*wronges* in (232)), adverbs (*paciently* in (232)), and auxiliaries (De Smet 2008: 56). De Smet (2008: 62) has shown that the frequency of the nominal type peaked before 1420, after which the clausal type began to take over. It may be valuable to test whether French loan verbs are implicated only in the rise of nominal gerunds (cf. (231)), or whether they played a role in the emergence of the clausal type as well (cf. (232)). One of the explanations for the development of the clausal gerund has been that English copied the French gerund (translated as *gérondif*), which has clausal features (Jack 1988), but Kranich (2006) has rejected this

explanation. Therefore — like for *do*-support and light verbs — it is unlikely that the rise of the gerund has been caused by French influx, as the gerund has developed from Old English deverbal nouns in *-ing* or *-ung* (Fischer, De Smet & Wurff 2017: 173), and its rise is presumably due to a number of factors (De Smet 2008). However, the high rates of French influx may have further boosted the ongoing expansion of the gerund construction. Apart from the gerund, French influx has also tentatively been proposed as a trigger for the development of *wh*-relatives (Mustanoja 1960: 192) and the loss of verb-second (Gerritsen 2010: 117; Haeberli 2010).

More generally, biases in loan word accommodation may also aid in describing inexplicable syntactic changes in other languages which have undergone intense contact. In fact, the findings in this thesis may just reveal the tip of the iceberg.

# Bibliography

Aboh, Enoch. 2022. Universal multilingualism: why we are born to code-mix. Conference presentation presented at the Advancing code-switching research workshop, Vigo.

Ackerman, Farrell & Gert Webelhuth. 1997. The Composition of (Dis)continuous Predicates: Lexical or Syntactic? *Acta Linguistica Hungarica* 44(3/4). 317–340.

Adalar, Nevin & Sali Tagliamonte. 1998. Borrowed Nouns; Bilingual People: The Case of the "Londrali" in Northern Cyprus. *International Journal of Bilingualism* 2(2). 139–159. https://doi.org/10.1177/136700699800200202.

Adamou, Evangelia & Xingjia Rachel Shen. 2019. There are no language switching costs when codeswitching is frequent. *International Journal of Bilingualism* 23(1). 53–70. https://doi.org/10.1177/1367006917709094.

Akimoto, Minoji & Laurel J. Brinton. 1999. The Origin of the Composite Predicate in Old English. In Laurel J. Brinton & Minoji Akimoto (eds.), *Collocational and Idiomatic Aspects of Composite Predicates in the History of English*, 21–88. Amsterdam/Philadelphia: John Benjamins Publishing Company. https://benjamins.com/catalog/slcs.47.11aki.

Algeo, John. 1995. Having a Look at the Expanded Predicate. In Bas Aarts & Charles F. Meyer (eds.), *The verb in contemporary English: theory and description*, 75–101. Cambridge, NY: Cambridge University Press.

Algeo, John. 2006. *British or American English?: A Handbook of Word and Grammar Patterns*. 1st edn. Cambridge University Press. https://doi.org/10.1017/CBO9780511607240.

Allen, Cynthia. 1997. Middle English case loss and the 'creolization' hypothesis. *English Language & Linguistics* 1(1). 63–89. https://doi.org/10.1017/S1360674300000368.

Allerton, David John. 2002. *Stretched Verb Constructions in English*. London: Routledge.

Alsina, Alex, Joan Bresnan & Peter Sells. 1997. *Complex predicates*. Stanford: CSLI Publications.

Amuzu, Evershed Kwasi. 2016. The Syntax of Adverbials in Ewe-English Codeswitching. *Journal of Cognitive Science* 17(1). 133–165.

Andersen, Gisle. 2017. A corpus study of pragmatic adaptation: The case of the Anglicism [*jobb*] in Norwegian. *Journal of Pragmatics* 113. 127–143. https://doi.org/10.1016/j.pragma.2016.12.015.

Andersen, Gisle & Anne-Line Graedler. 2020. Morphological borrowing from English to Norwegian: The enigmatic non-possessive *-s*. *Nordic Journal of Linguistics* 43(1). 3–31. https://doi.org/10.1017/S0332586520000037.

Anderson, Gregory David Shelton. 2006. *Auxiliary Verb Constructions*. Oxford/New York: Oxford University Press.

Anderson, Gregory David Shelton. 2011. Auxiliary Verb Constructions (and Other Complex Predicate Types): A Functional-Constructional Overview: Auxiliary Verb Constructions. *Language and Linguistics Compass* 5(11). 795–828. https://doi.org/10.1111/j.1749-818X.2011.00311.x.

Androutsopoulos, Jannis. 2013. English 'on top': discourse functions of English resources in the German mediascape. *Sociolinguistic Studies* 6(2). 209–238. https://doi.org/10.1558/sols.v6i2.209.

Attali, Arlette & Simone Monsonégo. 1997. L'emploi des adjectifs dans Le Mesnagier de Paris (1394): Les adjectifs de couleur. In Bernard Combettes & Simone Monsonégo (eds.), *Le Moyen Français: Philologie et linguistique, approches du texte et du discours. Actes du VIIIe Colloque international sur le moyen français*, 211–232. Nancy: Didier Érudition.

Auer, Peter. 1998. *Code-switching in conversation: language, interaction and identity*. London: Routledge.

Auwera, Johan van der & Inge Genee. 2002. English *do*: on the convergence of languages and linguists. *English Language and Linguistics* 6(2). 283–307. https://doi.org/10.1017/S1360674302000242.

Babcock, Charlotte Farrington. 1914. A Study of the Metrical Use of the Inflectional *E* in Middle English, with Particular Reference to Chaucer and Lydgate. *Publications of the Modern Language Association of America* 29(1). 59–92. https://doi.org/10.1632/457034.

Backus, Ad. 2005. Codeswitching and language change: One thing leads to another? *International Journal of Bilingualism* 9(3–4). 307–340. https://doi.org/10.1177/13670069050090030101.

Backus, Ad. 2014. Towards a Usage-Based Account of Language Change: Implications of Contact Linguistics for Linguistic Theory. In Robert Nicolaï (ed.), *Questioning Language Contact*, 91–118. BRILL. https://doi.org/10.1163/9789004279056_005.

Backus, Ad. 2015. A usage-based approach to code-switching: The need for reconciling structure and function. In Gerald Stell & Kofi Yakpo (eds.), *Code-switching Between Structural and Sociolinguistic Perspectives*, 19–38. Berlin/München/Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110346879.19.

Backus, Ad. 2021. Usage-based approaches. In Evangelia Adamou & Yaron Matras (eds.), *The Routledge Handbook of Language Contact*, 110–126. London/New York: Routledge.

Backus, Ad & Derya Demirçay. 2021. Intense Turkish-Dutch bilingualism leads to intense Turkish-Dutch mixing: A usage-based account of increasing integration of two typologically different languages. *Belgian Journal of Linguistics* 35(1). 13–33. https://doi.org/10.1075/bjl.00062.bac.

Backus, Ad & Margreet Dorleijn. 2009. Loan translations versus code-switching. In Barbara E. Bullock & Almeida Jacqueline Toribio (eds.), *The Cambridge Handbook of Linguistic Code-switching*, 75–94. Cambridge: Cambridge University Press.

Bækken, Bjørg. 2000. Inversion in Early Modern English. *English Studies* 81(5). 393–421. https://doi.org/10.1076/0013-838X(200009)81:5;1-8:FT393.

Bailey, Charles James & Karl Maroldt. 1977. The French lineage of English. In Jürgen Meisel (ed.), *Pidgins – Creoles – Languages in contact*, 21–53. Tübingen: Narr.

Baker, Mona. 1996. Corpus-based translation studies: The challenges that lie ahead. In Harold Somers (ed.), *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager* (Benjamins Translation Library), vol. 18, 175–186. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Bakker, Peter. 1997. Athematic morphology in Romani: the borrowing of a borrowing pattern. In Yaron Matras, Peter Bakker & Hristo Kyuchukov (eds.), *The Typology and Dialectology of Romani*, 1–21. Amsterdam: John Benjamins Publishing Company.

Bakker, Peter. 2005. Intertwining and Michif. Conference presentation presented at the Romancisation Worldwide Conference, University of Bremen.

Balam, Osmer, María del Carmen Parafita Couto & Hans Stadthagen-González. 2020. Bilingual verbs in three Spanish/English code-switching communities. *International Journal of Bilingualism* 24(5–6). 952–967. https://doi.org/10.1177/1367006920911449.

Bartlett, Robert. 2000. *England under the Norman and Angevin kings, 1075-1225* (The New Oxford History of England). Oxford/New York: Clarendon Press/Oxford University Press.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). https://doi.org/10.18637/jss.v067.i01.

Baugh, Albert C. & Thomas Cable. 2013. *A History of the English Language*. 6th edn. Boston: Pearson Education, Inc.

Beatty, Jackson. 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin* 91(2). 276–292. https://doi.org/10.1037/0033-2909.91.2.276.

Belazi, Hedi M., Edward J. Rubin & Almeida Jacqueline Toribio. 1994. Code Switching and X-Bar Theory: The Functional Head Constraint. *Linguistic Inquiry* 25(2). 221–237.

Bellamy, Kate & M. Carmen Parafita Couto. 2022. Gender assignment in mixed noun phrases: State of the art. In Dalila Ayoun (ed.), *Studies in Bilingualism*, vol. 63, 13–48. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/sibil.63.02bel.

Berg, Thomas. 2014. Boundary permeability: A parameter for linguistic typology. *Linguistic Typology* 18(3). 489–531. https://doi.org/10.1515/lingty-2014-0020.

Berglund-Barraza, Amy, Fenghua Tian, Chandramalika Basak & Julia L. Evans. 2019. Word Frequency Is Associated With Cognitive Effort During Verbal Working Memory: A Functional Near Infrared Spectroscopy (fNIRS) Study. *Frontiers in Human Neuroscience* 13(433). 1–15. https://doi.org/doi:10.3389/fnhum.2019.00433.

Berteloot, Amand & Nicoline van der Sijs. 2002. Dutch. In Manfred Görlach (ed.), *English in Europe*. 1st edn. Oxford: Oxford University Press.

Besset, Ryan M. 2017. Exploring the Phonological Integration of Lone Other-Language Nouns in the Spanish of Southern Arizona. *University of Pennsylvania Working Papers in Linguistics* 23(2). 31–39.

Bezinska, Yanka & Iva Novakova. 2010. Grammaticalisation et acquisition des constructions causatives en français et en bulgare. *CogniTextes* 5(Volume 5). https://doi.org/10.4000/cognitextes.346.

Bhaskararao, Peri. 1998. Gadaba. In Sanford B. Steever (ed.), *The Dravidian languages*, 328–355. London: Routledge.

Biber, Douglas & Bethany Gray. 2013. Being Specific about Historical Change: The Influence of Sub-Register. *Journal of English Linguistics* 41(2). 104–134. https://doi.org/10.1177/0075424212472509.

Bijeljac-Babic, Ranka, Agnes Biardeau & Jonathan Grainger. 1997. Masked orthographic priming in bilingual word recognition. *Memory & Cognition* 25(4). 447–457.

Blackings, Mairi & Nigel Fabb. 2003. *A Grammar of Ma'di*. Berlin/New York: De Gruyter Mouton.

Blumenfeld, Henrike K. & Viorica Marian. 2007. Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and Cognitive Processes* 22(5). 633–660. https://doi.org/10.1080/01690960601000746.

Boas, Hans C. & Marc Pierce. 2011. Lexical developments in Texas German. *Studies on German-Language Islands* 129–150.

Bobb, Susan C. & Zofia Wodniecka. 2013. Language switching in picture naming: What asymmetric switch costs (do not) tell us about inhibition in bilingual speech planning. *Journal of Cognitive Psychology* 25(5). 568–585. https://doi.org/10.1080/20445911.2013.792822.

Bodleian Libraries. 2007. The British National Corpus, version 3 (BNC XML Edition). http://www.natcorp.ox.ac.uk/.

Bonial, Claire Nicole. 2014. *Take a Look at This! Form, Function and Productivity of English Light Verb Constructions*. Colorado: University of Colorado PhD dissertation.

Booij, Geert. 2001. English as the Lingua Franca of Europe: A Dutch Perspective. *Lingua e Stile* 36(2). 347–357.

Boon, Ton den & Ruud Hendrickx (eds.). 2015. *Van Dale Groot woordenboek van de Nederlandse taal*. 15th edn. Utrecht/Antwerpen: Van Dale Uitgevers.

Boumans, Louis. 2007. The periphrastic bilingual verb construction as a marker of intense language contact. Evidence from Greek, Portuguese and Maghribian Arabic. In Harald Motzki & Everhard Ditters (eds.), *Approaches to Arabic Linguistics*, 291–311. BRILL. https://doi.org/10.1163/ej.9789004160156.i-762.79.

Bowern, Claire. 2008. The diachrony of complex predicates. *Diachronica* 25(2). 161–185. https://doi.org/10.1075/dia.25.2.03bow.

Bowern, Claire. 2013. Relatedness as a Factor in Language Contact. *Journal of Language Contact* 6(2). 411–432. https://doi.org/10.1163/19552629-00602010.

Bowern, Claire, Patience Epps, Russell Gray, Jane Hill, Keith Hunley, Patrick McConvell & Jason Zentz. 2011. Does Lateral Transmission Obscure Inheritance in Hunter-Gatherer Languages? *PLoS ONE* 6(9). e25195. https://doi.org/10.1371/journal.pone.0025195.

Braine, Martin D.S. & Melissa Bowerman. 1976. Children's First Word Combinations. *Monographs of the Society for Research in Child Development* 41(1). 1–104.

Brinton, Laurel J. 1996. Attitudes toward Increasing Segmentalization: Complex and Phrasal Verbs in English. *Journal of English Linguistics* 24(3). 186–205. https://doi.org/10.1177/007542429602400304.

Brinton, Laurel J. & Minoji Akimoto. 1999. Introduction. In Laurel J. Brinton & Minoji Akimoto (eds.), *Collocational and Idiomatic Aspects of Composite Predicates in the History of English*, vol. 47. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Brinton, Laurel J. & Leslie K. Arnovick. 2011. *The English Language. A Linguistic History.* Oxford: Oxford University Press.

Broersma, Mirjam. 2009. Triggered codeswitching between cognate languages. *Bilingualism: Language and Cognition* 12(4). 447–462. https://doi.org/10.1017/S1366728909990204.

Brunger, Karen & Alison Henry. 1998. Processing Load and the Extended Optional Infinitive Stage. 33(S1). 422–427. https://doi.org/10.3109/13682829809179462.

Brunner, Karl. 1963. *An Outline of Middle English Grammar.* (Trans.) Grahame K. W. Johnston. Cambridge, MA: Harvard University Press.

Brysbaert, Marc, Boris New & Emmanuel Keuleers. 2012. Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods* 44(4). 991–997. https://doi.org/10.3758/s13428-012-0190-4.

Budts, Sara. 2020a. *On periphrastic do and the modal auxiliaries: A connectionist approach to language change.* Antwerp: University of Antwerp PhD dissertation.

Budts, Sara. 2020b. A connectionist approach to analogy. On the modal meaning of periphrastic *do* in Early Modern English. *Corpus Linguistics and Linguistic Theory* 18(2). 1–28. https://doi.org/10.1515/cllt-2019-0080.

Burnley, David. 1992. Lexis and Semantics. In Norman Blake (ed.), *The Cambridge history of the English language*, vol. 2, 409–99. Cambridge: Cambridge University Press.

Burrow, John A. & Thorlac Turville-Petre. 1992. *A Book of Middle English*. Oxford UK and Cambridge, MA: Blackwell.

Bybee, Joan. 2006. From Usage to Grammar: The Mind's Response to Repetition. *Language* 82(4). 711–733. https://doi.org/10.1353/lan.2006.0186.

Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Bybee, Joan & Paul J. Hopper. 2001. *Frequency and the emergence of linguistic structure*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Cambridge Dictionary. 2022a. Inversion. *Inversion*. https://dictionary-cambridge-org.kuleuven.e-bronnen.be/grammar/british-grammar/inversion. (1 August, 2022).

Cambridge Dictionary. 2022b. Do. *Do*. https://dictionary-cambridge-org.kuleuven.e-bronnen.be/grammar/british-grammar/do. (29 August, 2022).

Cambridge Dictionary. 2022c. Bear. *Bear*. https://dictionary-cambridge-org.kuleuven.e-bronnen.be/dictionary/english/bear?q=to+bear. (29 August, 2022).

Cambridge Dictionary. 2022d. Give. *Give*. https://dictionary-cambridge-org.kuleuven.e-bronnen.be/dictionary/english/give. (29 August, 2022).

Cambridge Dictionary. 2022e. Have. *Have*. https://dictionary-cambridge-org.kuleuven.e-bronnen.be/dictionary/english/have. (29 August, 2022).

Cambridge Dictionary. 2022f. Make. *Make*. https://dictionary-cambridge-org.kuleuven.e-bronnen.be/dictionary/english/make. (29 August, 2022).

Cambridge Dictionary. 2022g. Put. *Put*. https://dictionary-cambridge-org.kuleuven.e-bronnen.be/dictionary/english/put. (29 August, 2022).

Cambridge Dictionary. 2022h. Set. *Set*. https://dictionary-cambridge-org.kuleuven.e-bronnen.be/dictionary/english/set. (29 August, 2022).

Cambridge Dictionary. 2022i. Take. *Take*. https://dictionary-cambridge-org.kuleuven.e-bronnen.be/dictionary/english/take. (29 August, 2022).

Campbell, Lyle. 1998. *Historical Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.

Caon, Luisella. 2002. Final -E and Spelling Habits in the Fifteenth-Century Versions of the Wife of Bath's Prologue. *English Studies* 83(4). 296–310. https://doi.org/10.1076/enst.83.4.296.8686.

Cattell, Ray. 1984. 'Light' verbs in English. In Ray Cattell (ed.), *Composite Predicates in English*, vol. 17, 1–18. http://dx.doi.org/10.1163/9789004373136.

Claridge, Claudia. 2000. *Multi-word Verbs in Early Modern English: A Corpus-based Study*. Amsterdam Atlanta, GA: Rodopi.

Clark, Eve V. 1993. *The lexicon in acquisition*. Cambridge UK; New York: Cambridge University Press.

Clyne, Michael. 2003. *Dynamics of Language Contact: English and Immigrant Languages*. 1st edn. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511606526.

Clyne, Michael G. 1980. Triggering and language processing. *Canadian Journal of Psychology/Revue canadienne de psychologie* 34(4). 400–406. https://doi.org/10.1037/h0081102.

Coetsem, Frans van. 1988. *Loan Phonology and the Two Transfer Types in Language Contact*. Dordrecht: Foris Publications.

Coetsem, Frans van. 2000. *A general and unified theory of the transmission process in language contact*. Heidelberg: Winter.

Comrie, Bernard. 2008. Inflectional morphology and language contact, with special reference to mixed languages. In Peter Siemund & Noemi Kintana (eds.), *Language Contact and Contact Languages*, vol. 7, 15–32. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Coppen, Peter-Arno, Walter Haeseryn & Folkert de Vriend. 2002. Algemene Nederlandse Spraakkunst. http://ans.ruhosting.nl/e-ans/index.html. (24 March, 2020).

Costa, Albert, Alfonso Caramazza & Nuria Sebastian-Galles. 2000. The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26(5). 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283.

Coussé, Evie. 2013. The grammaticalization of the have perfect in Dutch – a corpus study of contextual extension and semantic generalization. *Language Sciences* 36. 103–112. https://doi.org/10.1016/j.langsci.2012.04.001.

Cowie, Claire. 2017. Morphology. In Alexander Bergs & Laurel J. Brinton (eds.), *The history of English*, 47–67. Berlin/Boston: De Gruyter Mouton.

Crespo, Begoña. 2000. Historical Background of Multilingualism and its Impact on English. In David Andrew Trotter (ed.), *Multilingualism in Later Medieval Britain*, 23–35. Cambridge: D.S. Brewer.

Culbertson, Jennifer, Marieke Schouwstra & Simon Kirby. 2020. From the world to word order: Deriving biases in noun phrase order from statistical properties of the world. *Language* 96(3). 696–717. https://doi.org/10.1353/lan.2020.0045.

Culpeper, Jonathan. 2005. *History of English*. 2nd edn. London: Taylor & Francis Ltd.

Curme, George. 1983. *A grammar of the English Language. Vol. 2: Syntax, reprint*. Essex, CT: Verbatim.

Dalton-Puffer, Christiane. 1996. *The French influence on Middle English morphology: a corpus-based study of derivation*. Berlin: De Gruyter Mouton.

Danchev, Andrei. 2011. The evidence for analytic and synthetic developments in English. In Matti Rissanen, Ossi Ihalainen, Terttu Nevalainen & Irma Taavitsainen (eds.), *History of Englishes: New Methods and Interpretations in Historical Linguistics*, 25–41. Berlin/Boston: De Gruyter Mouton.

Davidse, Kristin & Hendrik De Smet. 2020. Diachronic corpora. In *The practical handbook of corpus linguistics*. Berlin: Springer.

Davies, Mark. 2001. Corpus del Español. https://www.corpusdelespanol.org/.

De Cecco, Francesco. 2017. False Friends and True Cognates: On Fundamental Freedoms, Fundamental Rights and Union Citizenship. In Mads Andenas, Tarjei Bekkedal & Luca Pantaleo (eds.), *The Reach of Free Movement*, 253–271. The Hague: T.M.C. Asser Press. https://doi.org/10.1007/978-94-6265-195-1_11.

De Keyser, Silke. 2014. *Do-support and the Constant Rate Hypothesis*. Leuven: University of Leuven MA thesis.

De Smet, Hendrik. 2008. Functional motivations in the development of nominal and verbal gerunds in Middle and Early Modern English. *English Language and Linguistics* 12(1). 55–102. https://doi.org/10.1017/S136067430700250X.

De Smet, Hendrik. 2012. The course of actualization. *Language* 88(3). 601–633. https://doi.org/10.1353/lan.2012.0056.

De Smet, Hendrik. 2014. De integratie van Engelse leenwerkwoorden in het Nederlands. In Freek Van de Velde, Hans Smessaert, Frank Van Eynde & Sara Verbrugge (eds.), *Patroon en argument: Een dubbelfeestbundel bij het emeritaat van William Van Belle en Joop van der Horst*, 75–87. Leuven: Leuven University Press.

De Smet, Hendrik. 2020. What predicts productivity? Theory meets individuals. *Cognitive Linguistics* 31(2). 251–278. https://doi.org/10.1515/cog-2019-0026.

De Smet, Hendrik & Marlieke Shaw. 2022. Missing link: Code-switches, borrowings and accommodation biases. Manuscript submitted for publication.

De Sutter, Gert. 2017. Het Nederlands Anno 2017: Verspreiding, juridische status en variationele structuur. In Gert De Sutter (ed.), *De vele gezichten van het Nederlands in Vlaanderen: een inleiding tot de variatietaalkunde*, 22–32. Leuven/Den Haag: Acco.

Deibel, Isabel. 2020. The contribution of grammar and lexicon to language switching costs: Examining contact-induced languages and their implications for theories of language representation. *Bilingualism: Language and Cognition* 23(5). 992–1007. https://doi.org/10.1017/S1366728919000865.

Dekeyser, Xavier. 1986. Romance loans in Middle English: a re-assessment. In Dieter Kastovsky & Aleksander Szwedek (eds.), *Linguistics across historical and geographical boundaries*, 253–265. Berlin: De Gruyter Mouton.

Demirçay, Derya. 2017. *Connected languages: Effects of intensifying contact between Turkish and Dutch.* Tilburg: University of Tilburg PhD dissertation.

Denison, David. 1985. The origins of periphrastic DO: Ellegård and Visser reconsidered. In Roger Eaton, Olga Fischer, Willem F. Koopman & Frederike van der Leek (eds.), *Papers from the 4th International Conference on English Historical Linguistics*, vol. 41, 45–60. Amsterdam: John Benjamins Publishing Company.

Denison, David. 1993. *English Historical Syntax: Verbal Constructions*. London: Longman.

Denison, David. 2003. Log(ist)ic and simplistic S-curves. In Raymond Hickey (ed.), *Motives for Language Change*, 54–70. Cambridge: Cambridge University Press.

Desai, Rutvik H., Wonil Choi & John M. Henderson. 2020. Word frequency effects in naturalistic reading. *Language, Cognition and Neuroscience* 35(5). 583–594. https://doi.org/10.1080/23273798.2018.1527376.

Deuchar, Margaret. 2005. Congruence and Welsh–English code-switching. *Bilingualism: Language and Cognition* 8(3). 255–269. https://doi.org/10.1017/S1366728905002294.

Deuchar, Margaret. 2020. Code-Switching in Linguistics: A Position Paper. *Languages* 5(22). https://doi.org/10.3390/languages5020022.

Deuchar, Margaret. 2022. Controversies in code switching research. Conference presentation presented at the Advancing code-switching research workshop, Vigo.

Deumert, Ana. 2005. The unbearable lightness of being bilingual: English–Afrikaans language contact in South Africa. *Language Sciences* 27(1). 113–135. https://doi.org/10.1016/j.langsci.2004.10.002.

Diessel, Holger. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25(2). 108–127. https://doi.org/10.1016/j.newideapsych.2007.02.002.

Dijkstra, Ton. 2005. Bilingual Visual Word Recognition and Lexical Access. In Judith F. Kroll & Antoinette Maria Birgitta de Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, 179–201. Oxford: Oxford University Press.

Divjak, Dagmar & Catherine Caldwell-Harris. 2015. Frequency and entrenchment. In Ewa Dąbrowska & Dagmar Divjak (eds.), *Handbook of Cognitive Linguistics*, 53–75. Berlin: De Gruyter Mouton.

Dixon, Robert M. W. 1992. *A New Approach to English Grammar, on Semantic Principles*. Oxford: Clarendon Press.

Doğruöz, A. Seza & Eline Zenner. 2013. Socio-cultural variation in borrowability constraints: Fixedness and conventionality in contact situations. Conference presentation presented at the International Cognitive Linguistics Conferences, Edmonton, Alberta, Canada.

Domingue, Nicole Z. 1977. Middle English: Another creole? *Journal of Creole Studies* 1. 89–100.

Duff, Edward Gordon. 2017. *William Caxton*. Project Gutenberg.

Durkin, Philip. 2015. *Borrowed Words. A History of Loanwords in English*. Oxford: Oxford University Press.

Dutch Language Institute. 2015. SoNaR-corpus (Version 1.2.1). http://hdl.handle.net/10032/tm-a2-h5.

Edwards, Alison & Robert Fuchs. 2018. A cluster analysis of attitudes to English in Germany and the Netherlands. *World Englishes* 37(4). 653–667. https://doi.org/10.1111/weng.12348.

Eisen, Elad. 2019. *The Typology of Phonological Segment Borrowing*. Jerusalem: The Hebrew University of Jerusalem MA thesis. http://rgdoi.net/10.13140/RG.2.2.36245.55529.

Elenbaas, Marion. 2013. The synchronic and diachronic status of English light verbs. *Linguistic Variation* 13(1). 48–80. https://doi.org/10.1075/lv.13.1.02ele.

Elfner, Emily. 2018. The syntax-prosody interface: current theoretical approaches and outstanding questions. *Linguistics Vanguard* 4(1). 1–14. https://doi.org/10.1515/lingvan-2016-0081.

Ellegård, Alvar. 1953. *The Auxiliary Do: The Establishment and Regulation of Its Use in English*. Stockholm: Almqvist & Wiksell.

Elter, Wiebke Juliane. 2020. The rise of the to-dative: a language-contact approach to a phenomenon of structural language change. *Mannheim papers in multilingualism, acquisition and change acquisition and change*, 1. 1–71. https://doi.org/10.25521/mapmac.2020.121.

Elter, Wiebke Juliane & Marlieke Shaw. 2022. Loan verb accommodation: A comparison of Old Norse and French in Middle English. Conference presentation presented at the International Conference on Historical Linguistics (ICHL25), Oxford, UK.

Emonds, Joseph Embley & Jan Terje Faarlund. 2014. *English: The Language of the Vikings*. 1st ed. Olomouc: Palacký University.

Encyclopædia Britannica. n.d. Nahuatl language. https://www-britannica-com.kuleuven.e-bronnen.be/topic/Nahuatl-language. (1 September, 2022).

Encyclopædia Britannica. n.d. House of Plantagenet. *Encyclopædia Britannica*. https://www-britannica-com.kuleuven.e-bronnen.be/topic/house-of-Plantagenet. (30 August, 2022).

Encyclopædia Britannica. n.d. macaronic. https://www-britannica-com.kuleuven.e-bronnen.be/art/macaronic. (25 September, 2022).

Encyclopædia Britannica. n.d. Pentameter. *Encyclopædia Britannica*. Britannica Academic. https://academic-eb-com.kuleuven.e-bronnen.be/levels/collegiate/article/pentameter/59116. (26 July, 2021).

European Union. 2006. Eurobarometer: Europeans and their languages. https://europa.eu/eurobarometer/screen/home. (25 September, 2022).

Everaert, Martin. 2008. Collocations in Dutch. *Lexicographica* 24(2008). 91–102. https://doi.org/10.1515/9783484605336.1.91.

Fanego, Teresa. 2004. On reanalysis and actualization in syntactic change: The rise and development of English verbal gerunds. *Diachronica* 21(1). 5–55. https://doi.org/10.1075/dia.21.1.03fan.

Farrar, Kimberley & Mari C. Jones. 2002. Introduction. In *Language Change: The Interplay of Internal, External and Extra-Linguistic Factors*, 1–16. The Hague: De Gruyter Mouton.

Field, Fredric W. 2002. *Linguistic Borrowing in Bilingual Contexts*. Amsterdam: John Benjamins Publishing Company.

Filppula, Markku & Juhani Klemola. 2010. Celtic Influence on English: A Re-Evaluation. In Robert A. Cloutier, Anne Marie Hamilton-Brehm & William A. Kretzschmar Jr. (eds.), *Studies in the History of the English Language V: Variation and Change in English Grammar and Lexicon: Contemporary Approaches*, vol. 68, 207–230. Berlin/New York: De Gruyter Mouton.

Finkenstaedt, Thomas & Dieter Wolff. 1973. *Ordered Profusion. Studies in Dictionaries and the English Lexicon*. Heidelberg: Carl Winter.

Fischer, Hanna. 2018. *Präteritumschwund im Deutschen: Dokumentation und Erklärung eines Verdrängungsprozesses*. Boston: De Gruyter Mouton.

Fischer, Olga. 2007. *Morphosyntactic change: functional and formal perspectives*. Oxford: Oxford University Press.

Fischer, Olga. 2013. The role of contact in English syntactic change in the Old and Middle English periods. In *English as a contact language*, 18–40. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511740060.002.

Fischer, Olga, Hendrik De Smet & Wim van der Wurff. 2017. *A brief history of English syntax*. Cambridge: Cambridge University Press.

Fischer, Olga & Wim van der Wurff. 2006. Syntax. In Richard Hogg & David Denison (eds.), *A History of the English Language*, 109–198. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511791154.004.

Flamenbaum, Rachel. 2014. The pragmatics of codeswitching on Ghanaian talk radio. *International Journal of Bilingualism* 18(4). 346–362. https://doi.org/10.1177/1367006913481136.

Fuchs, Catherine & Sylvie Garnier. 2021. Pour un apprentissage/enseignement des collocations verbo-nominales d'existence en termes de 'constructions.' *Lexis* (18). https://doi.org/10.4000/lexis.5773.

Gardani, Francesco. 2008. *Borrowing of inflectional morphemes in language contact*. Frankfurt am Main: Peter Lang.

Gardani, Francesco. 2018. On morphological borrowing: On morphological borrowing. *Language and Linguistics Compass* 12(10). 1–17. https://doi.org/10.1111/lnc3.12302.

Gardner-Chloros, Penelope. 2009. *Code-switching*. Cambridge: Cambridge University Press.

Garrett, Andrew. 1998. On the origin of auxiliary *do*. *English Language and Linguistics* 2(2). 283–330. https://doi.org/10.1017/S1360674300000897.

Gąsiorowski, Piotr. 1997. Words in *-ate* and the history of English stress. In Jacek Fisiak (ed.), *Studies in Middle English Linguistics*, 157–180. Berlin: De Gruyter Mouton.

Gerritsen, Marinel. 2010. Divergent word order developments in Germanic languages: A description and a tentative explanation. In Jacek Fisiak (ed.), *Historical Syntax*, 107–135. Berlin/New York: De Gruyter Mouton.

Görlach, Manfred. 2003. Syntax. In Manfred Görlach (ed.), *Introduction to early modern English*, 95–135. Cambridge: Cambridge University Press.

Gosselin, Leah. 2021. The phonological integration of borrowings and code-switches: Support for the minimalist view of the bilingual language faculty. Conference presentation presented at the PsychoShorts 2021 Conference, Ottawa, Canada.

Gosselin, Leah. 2022. Bilinguals have a single computational system but two compartmentalized phonological grammars: Evidence from code-switching. *Glossa: a journal of general linguistics* 6(1). 147. https://doi.org/doi: https://doi.org/10.16995/glossa.5800.

Gosselin, Leah & Gabrielle Manning. 2022. Psycholinguists beware! Switching and borrowing involve distinct processing mechanisms. https://doi.org/10.13140/RG.2.2.34891.05925.

Gotthard, Lisa. 2019. Why *Do*-Support in Scots is Different. *English Studies* 100(3). 314–338. https://doi.org/10.1080/0013838X.2018.1558699.

Gottlieb, Henrik. 2012. Phraseology in flux: Danish Anglicisms beneath the surface. In Cristiano Furiassi, Virginia Pulcini & Félix Rodríguez González (eds.), *Anglicization of European Lexis*, 169–198. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Graedler, Anne-Line. 2012. The collection of Anglicisms: Methodological issues in connection with impact studies in Norway. In Cristiano Furiassi, Virginia Pulcini & Félix Rodríguez González (eds.), *The Anglicization of European Lexis*, 91–109. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Graedler, Anne-Line. 2014. Attitudes towards English in Norway: A corpus-based study of attitudinal expressions in newspaper discourse. *Multilingua* 33(3–4). 291–312. https://doi.org/10.1515/multi-2014-0014.

Graedler, Anne-Line & Guðrún Kvaran. 2010. Foreign influence on the written language in the Nordic language communities. *International Journal of the Sociology of Language* 2010(204). 31–42. https://doi.org/10.1515/ijsl.2010.029.

Grainger, Jonathan & Ton Dijkstra. 1992. On the Representation and Use of Language Information in Bilinguals. *Advances in Psychology* 83, 207–220. Elsevier. https://doi.org/10.1016/S0166-4115(08)61496-X.

Green, Clarence. 2017. *Patterns and Development in the English Clause System: A Corpus-Based Grammatical Overview.* 1st edn. Singapore: Springer.

Green's Dictionary of Slang. 2022. Cha. *Green's Dictionary of Slang.* (6 September, 2022).

Grondelaers, Stefan, Steven Delarue & Gert De Sutter. 2017. Tussen droom en daad: De consequenties van de verander(en)de standaardtaalrealiteit voor de taaldocent. In Gert De Sutter (ed.), *De vele gezichten van het Nederlands in Vlaanderen: Een inleiding tot de variatietaalkunde*, 330–352. Leuven/Den Haag: Acco.

Groot, Annette M. B. de & Peter A. Starreveld. 2015. Parallel language activation in bilinguals' word production and its modulating factors: a review and computer simulations. In John W. Schwieter (ed.), *The Cambridge Handbook of Bilingual Processing*, 389–415. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107447257.018.

Gullberg, Marianne. 2022. How to study code-switching (in production): a multi-task approach. Conference presentation presented at the Advancing code-switching research workshop, Vigo.

Gumperz, John J. 1982. *Discourse Strategies.* Cambridge: Cambridge University Press.

Haeberli, Eric. 2010. Investigating Anglo-Norman Influence on Late Middle English Syntax. In Richard Ingham (ed.), *The Anglo-Norman Language and Its Context*, 143–163. York: York Medieval Press.

Haeberli, Eric & Tabea Ihsane. 2016. Revisiting the loss of verb movement in the history of English. *Natural Language & Linguistic Theory* 34(2). 497–542.

Hall, Charles. 1983. *Periphrastic do: history and hypotheses.* Florida: University of Florida PhD dissertation. https://ufdcimages.uflib.ufl.edu/AA/00/04/77/60/00001/periphrasticdohi00hall.pdf.

Hansen, Erik & Jørn Lund. 1994. *Kulturens Gesandter. Fremmedordene i dansk.* Copenhagen: Munksgaard.

Hantson, André. 2003. English gerund clauses and Norwegian det + infinitive / at clause constructions. In Sylviane Granger, Jacques Lerot & Stephanie Petch-Tyson (eds.), *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*, vol. 20, 75–90. Amsterdam/New York: Rodopi.

Hardman, Martha James, Juana Vásquez & Juan de Dios Yapita. 1988. *Aymara. Compendio de estructura fonológica y gramatical.* La Paz: Instituto de Lengua y Cultura Aymara.

Harley, Heidi. 2007. The bipartite structure of verbs cross-linguistically (or: Why Mary can't exhibit John her paintings). Conference presentation presented at the ABRALIN, Belo Horizonte. http://heidiharley.com/heidiharley/wp-content/uploads/2016/08/harley_07_The-bipartite-s.pdf.

Harris, Alice & Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.

Haspelmath, Martin. 2008. Loanword typology: Steps toward a systematic cross-linguistic study of lexical borrowability. In Thomas Stolz, Dik Bakker & Rosa Salos Palomo (eds.), *Aspects of language contact: New theoretical, methodological and empirical findings with special focus on Romancisation processes*, vol. 35, 43–62. Berlin/New York: De Gruyter Mouton.

Hassan, Najat Abdulrahman & Nuha F. Malallah. 2020. Translating English Light Verb Constructions vs. Their Full Verbs into Arabic. *Journal of Al-Ma'moon College* (35). 255–285.

Haugen, Einar. 1950. The Analysis of Linguistic Borrowing. *Language* 26(2). 210–231.

Heine, Bernd & Tania Kuteva. 2008. Constraints on Contact-Induced Linguistic Change. *Journal of Language Contact* 2(1). 57–90. https://doi.org/10.1163/000000008792525363.

Heuven, Walter J.B. van, Ton Dijkstra & Jonathan Grainger. 1998. Orthographic Neighborhood Effects in Bilingual Word Recognition. *Journal of Memory and Language* 39(3). 458–483. https://doi.org/10.1006/jmla.1998.2584.

Hock, Hans Henrich & Brian D. Joseph. 2019. *Language history, language change, and language relationship: an introduction to historical and comparative linguistics*. 3rd revised edn. Berlin/Boston: De Gruyter.

Hockett, Charles Francis. 1969. *A Course in Modern Linguistics*. 14th edn. New York: The Macmillan Company.

Hofmann, Heike. 2000. Exploring categorical data: interactive mosaic plots. *Metrika* 51(1). 11–26.

Hollmann, Willem B. 2020. The 'nouniness' of attributive adjectives and 'verbiness' of predicative adjectives: evidence from phonology. *English Language and Linguistics* 25(2). 257–279. https://doi.org/10.1017/S1360674320000015.

Hoshino, Noriko & Judith F. Kroll. 2008. Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition* 106(1). 501–511. https://doi.org/10.1016/j.cognition.2007.02.001.

Houser, Michael, Line Mikkelsen, Ange Strom-Weber & Maziar Toosarvandani. 2006. Gøre-Support in Danish. In *Proceedings of the Twenty-First Comparative Germanic Syntax Workshop*.

Hout, Roeland van & Pieter Muysken. 1994. Modeling lexical borrowability. *Language Variation and Change* 6(1). 39–62. https://doi.org/10.1017/S0954394500001575.

Huddleston, Rodney. 1976. Some theoretical issues in the description of the English verb. *Lingua* 40. 331–383.

Hug, Sibylle. 1987. *Scandinavian Loanwords and their Equivalents in Middle English*. Bern/Frankfurt am Main/New York: Peter Lang.

Hunt, Tony. 2008. Anglo-Norman and the Loss of Normandy. In Florence Bourgne, Leo Carruthers & Arlette Sancery (eds.), *Un Espace Colonial et ses Avatars: Naissance D'Identités Nationales: Angleterre, France, Irlande (Ve-XVe siècles)*, 141–152. Paris: Presses de l'Université Paris-Sorbonne.

Iglesias-Rábade, Luis. 2000. French Phrasal Power in Late Middle English: Some Evidence Concerning the Verb nime(n)/take(n). In David Andrew Trotter (ed.), *Multilingualism in Later Medieval Britain*, 93–130. Cambridge: D.S. Brewer.

Iglesias-Rábade, Luis. 2001. Composite Predicates in Middle English with the Verbs nimen and taken. *Studia Neophilologica* 73(2). 143–163. https://doi.org/10.1080/003932701753401456.

Ingham, Richard. 2006. Syntactic change in Anglo-Norman and continental French chronicles: was there a 'Middle' Anglo-Norman? *Journal of French Language Studies* 16(1). 25–49. https://doi.org/10.1017/S0959269506002274.

Ingham, Richard. 2009a. The Persistence of Anglo-Norman 1230–1362: A Linguistic Perspective. *Language and culture in medieval Britain. The French of England* 44–54.

Ingham, Richard. 2009b. Mixing Languages on the Manor. *Medium Ævum* 78(1). 80–97.

Ingham, Richard. 2011a. Grammar change in Anglo-Norman and Continental French: The replacement of non-affirmative indefinite *nul* by *aucun*. *Diachronica* 28(4). 441–467.

Ingham, Richard. 2011b. Code-switching in the later medieval English lay subsidy rolls. In Herbert Schendl & Laura Wright (eds.), *Code-Switching in Early English*, 95–114. Berlin/Boston: De Gruyter Mouton.

Ingham, Richard. 2012. *The Transmission of Anglo-Norman. Language history and language acquisition*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Ingham, Richard. 2020. How Contact with French Drove Patient-Lability in English. *Transactions of the Philological Society* 118(3). 447–467. https://doi.org/10.1111/1467-968X.12199.

Jack, George. 1988. The Origins of the English Gerund. *North-Western European Language Evolution* 12(1). 15–75. https://doi.org/10.1075/nowele.12.02jac.

Jäger, Andreas. 2004. The Cross-Linguistic Function of Obligatory 'do'-Periphrasis. In *Proceedings of the 2004 Conference of the Australian Linguistic Society*. Sydney. http://ses.library.usyd.edu.au/bitstream/2123/111/1/ALS-20050630 - AJ.pdf.

Jäger, Andreas. 2006. *Typology of periphrastic 'do'-constructions*. Vol. 12. Bochum: Brockmeyer.

Jake, Janice L. & Carol Myers-Scotton. 2002. Making a minimalist approach to codeswitching work: Adding the matrix language. *Bilingualism: Language and Cognition* 5(1). 69–91. https://doi.org/10.1017/S1366728902000147.

Janssens, Guy & Ann Marynissen. 2008. *Het Nederlands vroeger en nu*. Leuven/Den Haag: Acco.

Jeppson, Haley, Heike Hofmann, Di Cook & Hadley Wickman. 2021. *ggmosaic: Mosaic Plots in the "ggplot2" Framework. R package version 0.3.3*. https://CRAN.R-project.org/package=ggmosaic.

Jespersen, Otto. 1905. *Growth and Structure of the English Language*. Leipzig: Teubner Publisher.

Jespersen, Otto. 1949. *A Modern English Grammar on Historical Principles*. Vol. I. Copenhagen: Munksgaard.

Johanson, Lars. 2002. Contact-induced change in a code-copying framework. In Mari C. Jones & Edith Esch (eds.), *Language Change: The Interplay of Internal, External and Extra-Linguistic Factors*, vol. 86, 285–313. Berlin: De Gruyter Mouton.

Johanson, Lars. 2008. Remodeling grammar: Copying, conventionalization, grammaticalization. In Peter Siemund & Noemi Kintana (eds.), *Language Contact and Contact Languages*, vol. 7, 61–79. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Kabatek, Johannes, Philipp Obrist & Valentina Vincis. 2010. Clause-linkage techniques as a symptom of discourse traditions: methodological issues and evidence from Romance languages. In Heidrun Dorgeloh & Anja Wanner (eds.), *Syntactic Variation and Genre*, 247–275. Berlin/New York: De Gruyter Mouton. https://doi.org/10.5167/UZH-86156.

Kampen, Jacqueline van & Frank Wijnen. 2000. Grammaticale ontwikkeling. In Steven Gillis & Annemarie Schaerlaekens (eds.), *Kindertaalverwerving: een handboek voor het Nederlands*, 225–285. Groningen: Nijhoff.

Kauhanen, Henri & George Walkden. 2018. Deriving the Constant Rate Effect. *Natural Language & Linguistic Theory* 36(2). 483–521. https://doi.org/10.1007/s11049-017-9380-1.

Keller, Mareike L. 2020. *Code-Switching: Unifying Contemporary and Historical Perspectives*. Cham: Springer. https://doi.org/10.1007/978-3-030-34667-6.

Kemenade, Ans van. 1987. *Syntactic Case and Morphological Case in the History of English*. Utrecht: University of Utrecht PhD dissertation.

Kerswill, Paul & Ann Williams. 2002. "Salience" as an explanatory factor in language change: evidence from dialect levelling in urban England. In Mari C. Jones & Edith Esch (eds.), *Language Change: The Interplay of Internal, External and Extra-Linguistic Factors*, vol. 86, 81–110. Berlin: De Gruyter Mouton.

Keuleers, Emmanuel, Marc Brysbaert & Boris New. 2010. SUBTLEX-NL: A new frequency measure for Dutch words based on film subtitles. *Behavior Research Methods* 42(3). 643–650.

Kibbee, Douglas A. 1991. *For to speke Frenche trewely: The French language in England, 1000–1600*. Amsterdam: John Benjamins Publishing Company.

Klemola, Juhani. 1998. Semantics of DO in South-Western dialects of English. In Ingrid Tieken-Boon van Ostade, Marijke van der Wal & Arjan van Leuvensteijn (eds.), *Do in English, Dutch, and German: History and present-day variation*, 25–52. Münster: Nodus Publikationen.

Koenig, Jean-Pierre. 1999. On a tué le président! The nature of passives and ultra-indefinites. In Barbara Fox, Dan Jurafsky & Laura A. Michaelis (eds.), *Cognition and function in language*, 235–251. Stanford: CSLI Publications.

Kotzé, Ernst F. 2001. Adjectival inflection in Afrikaans diachronics: An argument against the validity of creolization checklists. *Journal of Germanic Linguistics* 13(4). 381–391. https://doi.org/10.1017/S1470542701036042.

Kranich, Svenja. 2006. The origin of English gerundial constructions: A case of French influence? In Andrew James Johnston, Ferdinand von Mengden & Stefan Thim (eds.), *Language and Text: Current Perspectives on English and German Historical Linguistics and Philology*, 179–195. Heidelberg: Universitätsverlag Winter.

Kranich, Svenja, Viktor Becher & Stefen Höder. 2011. A tentative typology of translation-induced language change. In Svenja Kranich, Viktor Becher, Stefen Höder & Juliane House (eds.), *Multilingual Discourse Production. Diachronic and Synchronic Perspectives*, 11–44. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Kristensson, Gillis. 1997. The dialects of Middle English. In *Language History and Linguistic Modelling: A Festschrift for Jacek Fisiak on his 60th Birthday*. Reprint 2010. Berlin/Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110820751.655.

Kroch, Anthony. 2020. Penn Parsed Corpora of Historical English LDC2020T16.

Kroch, Anthony S. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1. 199–244.

Kroch, Anthony, Beatrice Santorini & Lauren Delfs. 2004. The Penn-Helsinki Parsed Corpus of Early Modern English. CD-ROM, 1st edn., release 3. http://www.ling.upenn.edu/ppche/ppche-release-2016/PPCEME-RELEASE-3.

Kroch, Anthony & Ann Taylor. 2000. Penn-Helsinki Parsed Corpus of Middle English, 2nd edn. http://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-4/index.html.

Krug, Manfred. 2003. Frequency as a determinant in grammatical variation and change. In Günter Rohdenburg & Britta Mondorf (eds.), *Determinants of Grammatical Variation in English*, 7–67. Berlin: De Gruyter Mouton.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar*. Stanford: Stanford University Press.

Lass, Roger. 1987. *The Shape of English: Structure and History*. London/Melbourne: J.M. Dent & Sons Ltd.

Lasser, Ingeborg. 2002. The roots of root infinitives: remarks on infinitival main clauses in adult and child language. *Linguistics* 40(4). 767–796. https://doi.org/10.1515/ling.2002.030.

Levelt, Willem J. M., Herbert Schriefers, Dirk Vorberg, Antje S. Meyer, Thomas Pechmann & Jaap Havinga. 1991. The time course of lexical access in speech production: A study of picture naming. *Psychological Review* 98(1). 122–142. https://doi.org/10.1037/0033-295X.98.1.122.

Levshina, Natalia. 2015. *How to do Linguistics with R: Data exploration and statistical analysis*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Lewis, Geoffrey L. 1985. *Turkish Grammar*. Oxford: Oxford University Press.

Lewis, Robert E. 1952. Middle English Dictionary. In Frances McSparran (ed.), *Online edition in Middle English Compendium*. Ann Arbor: University of Michigan Press. http://quod.lib.umich.edu/m/middle-english-dictionary/.

Li, Chuchu, Victor S. Ferreira & Tamar H. Gollan. 2022. Language control after phrasal planning: Playing Whack-a-mole with language switch costs. *Journal of Memory and Language* 126. 104338. https://doi.org/10.1016/j.jml.2022.104338.

Li, Chuchu & Tamar H. Gollan. 2018. Cognates interfere with language selection but enhance monitoring in connected speech. *Memory & Cognition* 46(6). 923–939. https://doi.org/10.3758/s13421-018-0812-x.

Los, Bettelou. 2016. The rise of verb-second. *Nederlandse Taalkunde* 21(2). 265–273. https://doi.org/10.5117/NEDTAA2016.2.LOS.

Los, Bettelou, Corrien Blom, Geert Booij, Marion Elenbaas & Ans van Kemenade. 2012. *Morphosyntactic Change: A Comparative Study of Particles and Prefixes*. 1st edn. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511998447.

Lumiansky, Robert M. 2019. Geoffrey Chaucer. https://www.britannica.com/biography/Geoffrey-Chaucer. (18 December, 2019).

MacKenzie, Ian. 2012. Fair play to them: Proficiency in English and types of borrowing. In Cristiano Furiassi, Virginia Pulcini & Félix Rodríguez González (eds.), *The Anglicization of European Lexis*, 27–42. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Mahlangu, KS. 2016. Language contact and linguistic change: The case of Afrikaans and English influence on isiNdebele. *South African Journal of African Languages* 36(1). 25–31. https://doi.org/10.1080/02572117.2016.1186892.

Mair, Christian. 2019. English in the German-Speaking World: An Inevitable Presence. In Raymond Hickey (ed.), *English in the German-Speaking World*, 13–30. 1st edn. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108768924.002.

Marian, Viorica & Michael Spivey. 2003. Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics* 24(2). 173–193. https://doi.org/10.1017/S0142716403000092.

Matras, Yaron. 2005. The full extent of "Fusion": A test case for connectivity and language contact. In Thomas Bierschenk, Walter Bisang, Detlev Kreikenbom & Ursula Verhoeven (eds.), *Prozesse des Wandels in historischen Spannungsfeldern Nordostafrikas / Westasiens. Akten zum 2.Symposium des SFB 295, Mainz 15.10–17.10.2001*, 241–255. Würzburg: Ergon.

Matras, Yaron. 2007. The borrowability of grammatical categories. In Yaron Matras & Sakel Jeanette (eds.), *Grammatical borrowing in cross-linguistic perspective*, 31–74. Berlin: De Gruyter Mouton.

Matras, Yaron. 2009. *Language Contact*. Cambridge: Cambridge University Press.

Matras, Yaron. 2013. Languages in contact in a world marked by change and mobility: *Revue française de linguistique appliquée* Vol. XVIII(2). 7–13. https://doi.org/10.3917/rfla.182.0007.

Matras, Yaron & Jeanette Sakel (eds.). 2007. *Grammatical borrowing in cross-linguistic perspective*. New York: De Gruyter Mouton.

Matsumoto, Meiko. 1999. Composite Predicates in Middle English. In Laurel J. Brinton & Minoji Akimoto (eds.), *Collocational and Idiomatic Aspects of Composite Predicates in the History of English*, 59. Amsterdam/Philadelphia: John Benjamins Publishing Company. https://www.jbe-platform.com/content/books/9789027298751-24mat.

McSparran, Frances. 2000a. Online edition in Middle English Compendium. https://quod.lib.umich.edu/m/middle-english-dictionary.

McSparran, Frances. 2000b. Corpus of Middle English Prose and Verse.

McWhorter, John H. 2002. What happened to English? *Diachronica* 19(2). 217–272. https://doi.org/10.1075/dia.19.2.02wha.

Meerts, Madeleine. 2013. *Toch niet zo gauw gefikst. De integratie van Engelse leenwerkwoorden in het Nederlands*. Leuven: University of Leuven MA thesis.

Meillet, Antoine. 1921. *Linguistique historique et linguistique générale*. Vol. VIII. Paris: Édouard Champion.

Mel'čuk, Igor. 1998. Collocations and lexical functions. In Anthony Paul Cowie (ed.), *Phraseology: Theory, Analysis and Applications*, 23–53. Oxford: Oxford University Pressμ.

Meuter, Renata F.I. & Alan Allport. 1999. Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language* 40. 25–40.

Miller, D. Gary. 2012. *External Influences on English: From Its Beginnings to the Renaissance*. Oxford: Oxford University Press.

Miyamoto, Tadao. 1999. *The Light Verb Construction in Japanese: The Role of the Verbal Noun*. Philadelphia, PA: John Benjamins Publishing Company.

Mohanan, Tara. 2006. Grammatical Verbs (With Special Reference to Light Verbs). In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell Companion to Syntax*, 459–492. Malden, MA: Blackwell. https://doi.org/10.1002/9780470996591.ch31.

Montague, Richard. 1974. *Formal philosophy London*. London: Yale University Press.

Moore, Colette. 2017. In Alexander Bergs & Laurel J. Brinton (eds.), *The history of English*, 287–308. Berlin/Boston: De Gruyter Mouton.

Moralejo-Garate, Teresa. 2001. Composite Predicates and idiomatisation in Middle English: a Corpus-based approach. *Studia Anglica Posnaniensia: International Review of English Studies* 36. 171–187.

Moravcsik, J. M. Edith. 1975a. Verb borrowing. *Wiener Linguistische Gazette* 8. 3–30.

Moravcsik, J. M. Edith. 1975b. *Understanding Language: A Study of Theories of Language in Linguistics and in Philosophy*. Vol. 169. The Hague: De Gruyter Mouton.

Moretti, Lorenzo. 2021. *On Multiple Constructions and Multiple Factors in Language Change: The origin of auxiliary do*. Manchester: University of Manchester PhD dissertation.

Moretti, Lorenzo. Multiplicity in language change: The origin of auxiliary *do*. Manuscript submitted for publication.

Mossé, Fernand. 1991. *Handbook of Middle English*. 10th edn. Baltimore, MD: The John Hopkins University Press.

Mugglestone, Lynda (ed.). 2006. *The Oxford History of English*. Oxford/New York: Oxford University Press.

Mustanoja, Tauno F. 1960. *A Middle English syntax. 1: Parts of speech*. Helsinki: Société néophilologique.

Muysken, Pieter Cornelis. 2013. Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition* 16(4). 709–730. https://doi.org/10.1017/S1366728912000727.

Muysken, Pieter Cornelis. 1981. Halfway between Quechua and Spanish: the case for relexification. In Arnold Highfield & Albert Valdman (eds.), *Historicity and variation in creole studies*, 52–78. Ann Arbor: Karoma Publishers.

Muysken, Pieter Cornelis. 2000. *Bilingual Speech: A Typology of Code-Mixing*. Cambridge: Cambridge University Press.

Myers-Scotton, Carol. 1993a. *Duelling languages: grammatical structure in codeswitching*. Oxford: Clarendon Press.

Myers-Scotton, Carol. 1993b. *Social Motivations For Codeswitching: Evidence from Africa*. Oxford: Clarendon Press.

Myers-Scotton, Carol. 2002. *Contact linguistics: bilingual encounters and grammatical outcomes*. Oxford/New York: Oxford University Press.

Myers-Scotton, Carol & Janice L. Jake. 2014. Nonfinite verbs and negotiating bilingualism in codeswitching: Implications for a language production model. *Bilingualism: Language and Cognition* 17(3). 511–525. https://doi.org/10.1017/S1366728913000758.

Naro, Anthony J. 1981. The social and structural dimensions of a syntactic change. *Language* 57(1). 63–98. https://doi.org/10.1353/lan.1981.0020.

Nash, Léa & Pollet Samvelian (eds.). 2016. *Approaches to complex predicates*. Leiden/Boston: BRILL.

Nederlandse Taalunie. 2004. Het project Corpus Gesproken Nederlands. https://ivdnt.org/images/stories/producten/documentatie/cgn_website/doc_Dutch/topics/index.htm. (14 November, 2019).

Nelson, Katherine. 1976. Some attributes of adjectives used by young children. *Cognition* 4. 13–30.

Nevalainen, Terttu, Helena Raumolin-Brunberg, Jukka Keränen, Minna Nevala, Arja Nurmi & Minna Palander-Collin. 1998. Corpus of Early English Correspondence. https://varieng.helsinki.fi/CoRD/corpora/CEEC/.

Nevalainen, Terttu, Tanja Säily, Turo Vartiainen, Aatu Liimatta & Jefrey Lijffijt. 2020. History of English as punctuated equilibria? A meta-analysis of the rate of linguistic change in Middle English. *Journal of Historical Sociolinguistics* 6(2). 20190008. https://doi.org/10.1515/jhsl-2019-0008.

Nickel, Gerhard. 1968. Complex Verbal Structures in English. *International Review of Applied Linguistics* 6(1). 63–83.

Ninio, Anat. 2004. Young children's difficulty with adjectives modifying nouns. *Journal of Child Language* 31(2). 255–285. https://doi.org/10.1017/S0305000904006191.

Nurmi, Arja. 2000. The rise and regulation of periphrastic do in negative declarative sentences: a sociolinguistic study. In Dieter Kastovsky & Arthur Mettinger (eds.), *The History of English in a Social Context*, 339–362. Berlin/New York: De Gruyter Mouton. https://doi.org/10.1515/9783110810301.339.

Nurmi, Arja. 2017. Early Modern English: Overview. In Alexander Bergs & Laurel J. Brinton (eds.), *The history of English*, 8–26. Berlin/Boston: De Gruyter Mouton.

Nuzzo, Regina L. 2021. Mosaic plots for displaying and exploring categorical data. *PM&R* 13(5). 531–534. https://doi.org/10.1002/pmrj.12593.

Onions, Charles Talbut, George Washington Salisbury Friedrichsen & Robert William Burchfield (eds.). 1966. *The Oxford dictionary of English etymology*. Oxford: Clarendon Press.

Onysko, Alexander. 2007. *Anglicisms in German: borrowing, lexical productivity, and written codeswitching*. Berlin/New York: Walter de Gruyter.

Onze Taal. 2022. Buigings-e: een groot / grote man. *Buigings-e: een groot / grote man*. https://onzetaal.nl/taaladvies/een-groot-grote-man. (7 April, 2022).

O'Shannessy, Carmel. 2015. Typological and social factors influencing a new mixed language, Light Warlpiri. In Gerald Stell & Kofi Yakpo (eds.), *Code-switching between structural and sociolinguistic perspectives*, vol. 43, 289–304. Berlin/Boston: De Gruyter Mouton.

Oxford University Press. 2009. Oxford English Dictionary Second Edition on CD-ROM (v. 4.0).

Parafita Couto, M. Carmen, Miriam Greidanus Romaneli & Kate Bellamy. Code-switching at the interface between language, culture, and cognition. Manuscript in press. https://hal.archives-ouvertes.fr/halshs-03280922/.

Parafita Couto, Maria Carmen & Marianne Gullberg. 2019. Code-switching within the noun phrase: Evidence from three corpora. *International Journal of Bilingualism* 23(2). 695–714. https://doi.org/10.1177/1367006917729543.

Percillier, Michael & Carola Trips. 2020. Lemmatising Verbs in Middle English Corpora: The Benefit of Enriching the Penn-Helsinki Parsed Corpus of Middle English 2 (PPCME2), the Parsed Corpus of Middle English Poetry (PCMEP), and A Parsed Linguistic Atlas of Early Middle English (PLAEME). In *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, 7170–7178. Marseille.

Petré, Peter, Lynn Anthonissen, Sara Budts, Enrique Manjavacas, Emma-Louise Silva, William Standing & Odile A.O. Strik. 2019. Early Modern Multiloquent Authors (EMMA): Designing a large-scale corpus of individuals' languages. *ICAME Journal* 43(1). 83–122. https://doi.org/10.2478/icame-2019-0004.

Phillips, Colin. 2010. Syntax at Age Two: Cross-Linguistic Differences. *Language Acquisition* 17(1/2). 70–120. https://doi.org/10.1080/10489221003621167.

Platzack, Christer. 2008. Cross Linguistic Variation in the Realm of Support Verbs. In *Proceedings of Comparative Germanic Syntax Workshop 23*. https://ling.auf.net/lingbuzz/000766.

Pol, Nikki van de. 2016. *The development of the Absolute Construction in English: Between bird's eye view and magnifying glass*. Leuven: University of Leuven PhD dissertation.

Poplack, Shana. 1980. "Sometimes I'll start a sentence in Spanish y termino en español": toward a typology of code-switching. *Linguistics* 18(7/8). 581–618.

Poplack, Shana. 2000. Introduction to reprinted version of Poplack (1980). In Li Wei (ed.), *The bilingualism reader*, 221–222. London: Routledge.

Poplack, Shana. 2018. *Borrowing: loanwords in the speech community and in the grammar*. New York: Oxford University Press.

Poplack, Shana & Nathalie Dion. 2012. Myths and facts about loanword development. *Language Variation and Change* 24(2012). 279–315. https://doi.org/10.1017/S095439451200018X.

Poplack, Shana & Marjory Meechan. 1999. The nature of constraints on contact-induced change. *Bilingualism: Language and Cognition* 2(2). 84–85. https://doi.org/10.1017/S1366728999230209.

Poplack, Shana, Suzanne Robillard, Nathalie Dion & John C. Paolillo. 2020. Revisiting Phonetic Integration in Bilingual Borrowing. *Language (Baltimore)* 96(1). 126–159. https://doi.org/10.1353/lan.2020.0004.

Poplack, Shana, David Sankoff & Christopher Miller. 1988. The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics* 26(1). 47–104. https://doi.org/10.1515/ling.1988.26.1.47.

Poplack, Shana, Susan Wheeler & Anneli Westwood. 1989. Distinguishing language contact phenomena: evidence from Finnish-English bilingualism. *World Englishes* 8(3). 389–406. https://doi.org/10.1111/j.1467-971X.1989.tb00677.x.

Poussa, Patricia. 1990. A contact-universals origins for periphrastic *do* with special consideration of old English-Celtic contact. In Sylvia M. Adamson, Vivien A. Law, Nigel Vincent & Susan Wright (eds.), *Current Issues in Linguistic Theory*, vol. 65, 407f. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/cilt.65.23pou.

Preussler, Walter. 1956. Keltischer Einfluss im Englischen. *Revue des Langues Vivantes* 22. 322–350.

Prins, Anton Adriaan. 1948. French Phrases in English. *Neophilologus* 32(1). 28–39.

Prins, Anton Adriaan. 1952. *French Influence in English Phrasing*. Leiden: Leiden University Press.

Pyles, Thomas & John Algeo. 1982. *The origins and development of the English language*. 3rd ed. New York: Harcourt Brace Jovanovich.

Quirk, Randolph, Sidney Greenbaum & Geoffrey Leech. 1985. *A Comprehensive Grammar of the English Language*. 1st edn. London: Longman.

Quirk, Randolph & Joan Mulholland. 1964. Complex prepositions and related sequences. *English Studies* 45(1–6). 64–73. https://doi.org/10.1080/00138386408597188.

Raichlin, Rina, Joel Walters & Carmit Altman. 2018. Some wheres and whys in bilingual codeswitching: Directionality, motivation and locus of codeswitching in Russian-Hebrew bilingual children. *International Journal of Bilingualism* 23(2). 629–650. https://doi.org/10.1177/1367006918763135.

Renault-Lescure, Odile. 2004. Contacts de langues et réanalyse de certaines structures prédicatives en kali'na (langue caribe). Conference presentation presented at the International Joint Conference of the SCL-SCPL-ACBLPE on Caribbean and Creole Languages, Curaçao.

Rey, Alain. 2016. *Dictionnaire historique de la langue française: L'origine et l'histoire des mots*. Vol. 4. Paris: Le Robert.

Rey Quesada, Santiago del. 2017. (Anti-)Latinate syntax in Renaissance dialogue: Romance translations of Erasmus's Uxor Mempsigamos. *Zeitschrift für romanische Philologie* 133(3). 673–708. https://doi.org/10.1515/zrp-2017-0034.

Rice, Mabel L., Karen Ruff Noll & Hannelore Grimm. 1997. An Extended Optional Infinitive Stage in German-Speaking Children With Specific Language Impairment. *Language Acquisition* 6(4). 255–295. https://doi.org/10.1207/s15327817la0604_1.

Rice, Mabel L. & Kenneth Wexler. 1996. Toward Tense as a Clinical Marker of Specific Language Impairment in English-Speaking Children. *Journal of Speech, Language, and Hearing Research* 39(6). 1239–1257. https://doi.org/10.1044/jshr.3906.1239.

Rissanen, Matti. 1985. Periphrastic *Do* in Affirmative Statements in Early American English. *Journal of English Linguistics* 18(2). 163–183. https://doi.org/10.1177/007542428501800205.

Rissanen, Matti. 1991. Spoken language and the history of *do*-periphrasis. In Dieter Kastovsky (ed.), *Historical English Syntax*, 321–342. Berlin/Boston: De Gruyter Mouton.

Rissanen, Matti. 2000. Syntax. In Roger Lass (ed.), *The Cambridge History of the English Language*, 187–331. 1st edn. Cambridge: Cambridge University Press. https://doi.org/10.1017/CHOL9780521264761.005.

Rissanen, Matti, Merja Kytö, Leena Kahlas-Tarkka, Matti Kilpiö, Saara Nevanlinna, Irma Taavitsainen, Terttu Nevalainen & Helena Raumolin-Brunberg. 1991. The Helsinki Corpus of English Texts. https://varieng.helsinki.fi/CoRD/corpora/HelsinkiCorpus/. (4 October, 2018).

Rohde, Ada, Anatol Stefanowitsch & Suzanne Kemmer. 2000. *Loanwords in a Usage-Based Model*. LAUD.

Romaine, Suzanne. 1989. *Bilingualism*. Oxford: Basil Blackwell.

Ronan, Patricia. 2012. Mobilizing linguistic concepts: support verb structures in early English. *E-Periodica*. https://doi.org/10.5169/SEALS-323530.

Ronan, Patricia. 2014. Light verb constructions in the history of English. In Kristin Davidse, Caroline Gentens, Lobke Ghesquière & Lieven Vandelanotte (eds.), *Studies in Corpus Linguistics*, vol. 63, 15–34. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/scl.63.05ron.

Rosemeyer, Malte & Scott A. Schwenter. 2019. Entrenchment and persistence in language change: the Spanish past subjunctive. *Corpus Linguistics and Linguistic Theory* 15(1). 167–204. https://doi.org/10.1515/cllt-2016-0047.

Rot, Sándor. 1991. *Language contact*. Frankfurt am Main/New York: Peter Lang.

Rothwell, William. 1968. The Teaching of French in Medieval England. *The Modern Language Review* 63(1). 37–46.

Rothwell, William. 1976. The role of French in thirteenth-century England. *Bulletin of the John Rylands Library* 58(2). 445–466. https://doi.org/10.7227/BJRL.58.2.8.

Rothwell, William. 1980. Lexical borrowing in a medieval context. *Bulletin of the John Rylands Library* 63. 118–43.

Rothwell, William. 1983. Language and Government in Medieval England. *Zeitschrift für französische Sprache und Literatur* 93(9). 258–270.

Rothwell, William. 1998. Arrivals and departures: The adoption of French terminology into middle English. *English Studies* 79(2). 144–165. https://doi.org/10.1080/00138389808599121.

Rothwell, William. 2000. Aspects of Lexical and Morphosyntactical Mixing. In David Andrew Trotter (ed.), *Multilingualism in Later Medieval Britain*, 213–232. Cambridge: D.S. Brewer.

Rothwell, William. 2001. English and French in England after 1362. *English Studies* 82(6). 539–559. https://doi.org/10.1076/enst.82.6.539.9550.

Samvelian, Pollet, Laurence Danlos & Benoît Sagot. 2011. On the predictability of light verbs. In *30th International Conference on Lexis and Grammar*. Nicosia, Cyprus. https://hal.inria.fr/inria-00617506.

Sankoff, David. 1998. A formal production-based explanation of the facts of code-switching. *Bilingualism: Language and Cognition* 1(1). 39–50. https://doi.org/10.1017/S136672899800011X.

Sankoff, David, Shana Poplack & Swathi Vanniarajan. 1990. The case of the nonce loan in Tamil. *Language Variation and Change* 2. 71–101.

Sankoff, Gillian. 2001. Linguistic Outcomes of Language Contact. In J.K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, 638–668. Oxford: Basil Blackwell. https://doi.org/10.1002/9780470756591.ch25.

Saugera, Valérie. 2012. The inflectional behavior of English-origin adjectives in French. *Journal of French Language Studies* 22(2). 225–250. https://doi.org/10.1017/S0959269511000032.

Schauwecker, Yela. 2018. Contact influence from Early English: preverbal elements in Anglo-French. Conference presentation presented at the Summer school in Historical Linguistics of Germanic Languages SUM-UP, Potsdam, Germany.

Schauwecker, Yela. 2019. Le faus françeis d'Angleterre en tant que langue seconde ? Quelques phénomènes syntaxiques indicatifs. *Revue des langues romanes* (Tome CXXIII N°1). 45–68. https://doi.org/10.4000/rlr.1488.

Schauwecker, Yela & Carola Trips. 2018. Who came riding first? Le chevalier or the knight? A multiple corpus analysis investigating historical language contact. In *CRH-2 : proceedings of the Second Workshop on Corpus-based Research in the Humanities 25-26 January 2018 Vienna, Austria*, 181–190. Vienna (Austria): Frank, Andrew U.

Schellinck, Leen. 2011. *De integratie van Engelse adjectieven en werkwoorden in het Nederlands*. Leuven: University of Leuven MA thesis.

Schendl, Herbert. 2000. Linguistic Aspects of Code-Switching in Medieval English Texts. In David Andrew Trotter (ed.), *Multilingualism in Later Medieval Britain*, 77–92. Cambridge: D.S. Brewer.

Schendl, Herbert. 2004. English historical code-switching in a European perspective. In J. Normann Jørgensen & Christine B. Dabelsteen (eds.), *Languaging and Language practices*, 188–202. Copenhagen: University of Copenhagen.

Schendl, Herbert. 2012. Middle English: Language contact. In Alexander Bergs & Laurel J. Brinton (eds.), *Historical linguistics of English*, 505–519. Berlin/New York: De Gruyter Mouton.

Schendl, Herbert & Laura Wright. 2011. *Code-Switching in Early English*. Berlin/Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110253368.

Scherling, Johannes. 2013. Holistic loanword integration and loanword acceptance: A comparative study of anglicisms in German and Japanese. *Arbeiten aus Anglistik und Amerikanistik* 38(1). 37–51.

Schlyter, Birgit N. 2003. Sociolinguistic changes in transformed Central Asian societies. In Jacques Maurais & Michael A. Morris (eds.), *Languages in a Globalising World*, 157–187. 1st edn. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511613739.011.

Schmid, Hans-Jörg. 2013. Is usage more than usage after all? The case of English not that. *Linguistics* 51(1). https://doi.org/10.1515/ling-2013-0003.

Schmid, Hans-Jörg. 2015. A blueprint of the Entrenchment-and- Conventionalization Model. *Yearbook of the German Cognitive Linguistics Association* 3(1). 3–25. https://doi.org/10.1515/gcla-2015-0002.

Schmid, Hans-Jörg. 2016. *Entrenchment and the Psychology of Language Learning: How We Reor-ganize and Adapt Linguistic Knowledge*. Berlin/Boston: De Gruyter Mouton.

Schmid, Hans-Jörg. 2020. *The dynamics of the linguistic system: usage, conventionalization, and entrenchment*. 1st edn. Oxford/New York: Oxford University Press.

Schmid, Hans-Jörg & Annette Mantlik. 2015. Entrenchment in Historical Corpora? Reconstructing Dead Authors' Minds from their Usage Profiles. *Anglia* 133(4). 583–623. https://doi.org/10.1515/anglia-2015-0056.

Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the world* (Cambridge Approaches to Language). Cambridge: Cambridge University Press.

Schultze-Berndt, Eva. 2017. Two classes of verbs in Northern Australian languages: Implications for the typology of polycategoriality. In Valentina Vapnarsky & Edy Veneziano (eds.), *Lexical Polycategoriality: Cross-linguistic, cross-theoretical and language acquisition approaches*, vol. 182, 243–70. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Schwenter, Scott A. & Elizabeth Close Traugott. 2000. Invoking scalarity: The development of in fact. *Journal of Historical Pragmatics* 1(1). 7–25.

Seifart, Frank. 2017. Patterns of affix borrowing in a sample of 100 languages. *Journal of Historical Linguistics* 7(3). 389–431. https://doi.org/10.1075/jhl.16002.sei.

Seifart, Frank. 2019. 2. Contact-induced change. In Jeroen Darquennes, Joseph C. Salmons & Wim Vandenbussche (eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK)*, 13–23. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110435351-002.

Shahrokny-Prehn, Arian & Silke Höche. 2011. Rising through the registers – A corpus-based account of the stylistic constraints on Light Verb Constructions. *Corpus* (10). 239–257. https://doi.org/10.4000/corpus.2110.

Shaw, Marlieke & Hendrik De Smet. 2022. Loan Word Accommodation Biases: Markedness and Finiteness. *Transactions of the Philological Society* 120(2). 201–217. https://doi.org/10.1111/1467-968X.12233.

Shaw, Marlieke & Hendrik De Smet. Predicative and markedness bias in loan adjectives: Dutch and Middle English. Manuscript in press.

Shechter, Adi & David L. Share. 2021. Keeping an Eye on Effort: A Pupillometric Investigation of Effort and Effortlessness in Visual Word Recognition. *Psychological Science* 32(1). 80–95. https://doi.org/10.1177/0956797620958638.

Short, Ian. 1980. Bilingualism in Anglo-Norman England. *Romance Philology* 33(4). 467–479.

Short, Ian. 2007. *Manual of Anglo-Norman*. London: Anglo-Norman text society.

Sijs, Nicoline van der. 1996. *Leenwoordenboek: de invloed van andere talen op het Nederlands*. Den Haag: Sdu Uitgevers.

Sijs, Nicoline van der. 2005. *Groot leenwoordenboek*. Utrecht: Van Dale Lexicografie.

Sijs, Nicoline van der. 2012. Engelse leenwoorden revisited. *Onze Taal* 81(5). 132–134.

Smans, Dorien. 2011. *De integratie van Engelse adjectieven in het Nederlands*. Leuven: University of Leuven MA thesis.

Smitterberg, Erik & Merja Kytö. 2015. English Genres in Diachronic Corpus Linguistics. In Philip Shaw, Britt Erman, Gunnel Melchers & Peter Sundkvist (eds.), *From Clerks to Corpora: essays on the English language yesterday and today*, 117–133. Stockholm: Stockholm University Press. https://doi.org/10.16993/bab.g.

Spencer, Andrew. 1991. *Morphological theory: an introduction to word structure in generative grammar*. Oxford, MS/Cambridge, MA: Basil Blackwell.

Stammers, Jonathan R. & Margaret Deuchar. 2012. Testing the nonce borrowing hypothesis: Counter-evidence from English-origin verbs in Welsh. *Bilingualism: Language and Cognition* 15(3). 630–643. https://doi.org/10.1017/S1366728911000381.

Stefanowitsch, Anatol. 2002. Nice to miet you: bilingual puns and the status of English in Germany. *Intercultural Communication Studies* 11. 67–84.

Stein, Achim & Carola Trips. 2012. Diachronic aspects of borrowing aspect: the role of Old French in the development of the be going to + INF construction. *SHS Web of Conferences* 1. 227–246. https://doi.org/10.1051/shsconf/20120100254.

Stein, Achim & Carola Trips. 2016. Borrowing of Argument Structure in Contact Situations (BASICS). *Borrowing of Argument Structure in Contact Situations (BASICS).* https://sites.google.com/site/dfgbasics/home. (15 September, 2022).

Stein, Dieter. 1990. *The Semantics of Syntactic Change: Aspects of the Evolution of "do" in English*. Vol. 47. Berlin/Boston: De Gruyter Mouton.

Stein, Dieter. 2011. *Do* up to the fifteenth century. In Dieter Stein (ed.), *The Semantics of Syntactic Change: Aspects of the Evolution of "do" in English*, 13–30. Berlin/New York: De Gruyter Mouton. https://doi.org/10.1515/9783110846829.

Stene, Aasta. 1945. *English Loan-words in Modern Norwegian: A Study of Linguistic Borrowing in the Process*. University of Michigan: Philological Society.

Stern, Henry R. 1977. English in Flemish Belgium. *American Speech* 52(1–2). 128–133. https://doi.org/10.2307/454727.

Strang, Barbara M. H. 1970. *A History of English*. London/Colchester: Methuen & Co Ltd.

Swaan, Abram de. 2001. *Words of the World*. Cambridge: Polity Press.

Swaan, Abram de. 2002. *The World Language System: A Political Sociology and Political Economy of Language*. Cambridge: Polity Press.

Swaan, Abram de. 2016. Dutch in the World Language System. *CLINA: Revista Interdisciplinaria de Traducción, Interpretación y Comunicación Intercultural* 2(2). 11–14. https://doi.org/10.14201/clina2016221114.

Swinburne, Nicola. 2022. 'Do'-support in the northern Italian Camuno. *Proceedings of the Linguistic Society of America* 7(1). 5211. https://doi.org/10.3765/plsa.v7i1.5211.

Tanabe, Harumi. 1999. Composite Predicates and Phrasal Verbs in The Paston Letters. In Laurel J. Brinton & Minoji Akimoto (eds.), *Collocational and Idiomatic Aspects of Composite Predicates in the History of English*, 97–132. Amsterdam/Philadelphia: John Benjamins Publishing Company. https://benjamins.com/catalog/slcs.47.36tan.

Taylor, Ann. 2008. Contact effects of translation: Distinguishing two kinds of influence in Old English. *Language Variation and Change* 20(2). 341–365. https://doi.org/10.1017/S0954394508000100.

Taylor, Ann, Arja Nurmi, Anthony Warner, Susan Pintzuk, Terttu Nevalainen & CEEC Project Team. 2006. Parsed Corpus of Early English Correspondence, parsed version. York: University of York and Helsinki: University of Helsinki. Distributed through the Oxford Text Archive.

Text Creation Partnership. 2014. Early English Books Online. https://www-english-corpora-org.kuleuven.e-bronnen.be/eebo/.

Thomason, Sarah Grey. 2016. Middle English: English, Not Norse. *Language Dynamics and Change* 6(1). 42–45. https://doi.org/10.1163/22105832-00601010.

Thomason, Sarah Grey. 2001. *Language Contact. An Introduction*. Edinburgh: Edinburgh University Press.

Thomason, Sarah Grey. 2003. Contact as a Source of Language Change. In Janda Richard D. & Joseph Brian D. (eds.), *A handbook of historical linguistics*, 687–712. Oxford: Blackwell.

Thomason, Sarah Grey & Terrence Kaufman. 1991. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Tieken-Boon van Ostade, Ingrid. 1990. The Origin and Development of Periphrastic Auxiliary Do: A Case of Destigmatisation. *North-Western European Language Evolution* 16. 3–52. https://doi.org/10.1075/nowele.16.01ost.

Timofeeva, Olga & Richard Ingham. 2018. Special issue on mechanisms of French contact influence in Middle English: diffusion and maintenance. *English Language and Linguistics* 22(2). 197–205. https://doi.org/10.1017/S1360674318000096.

Tomasello, Michael. 1992. *First Verbs: A Case Study of Early Grammatical Development*. 1st edn. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511527678.

Torres Cacoullos, Rena. 2020. Code-Switching Strategies: Prosody and Syntax. *Frontiers in Psychology* 11. 2130. https://doi.org/10.3389/fpsyg.2020.02130.

Torres Cacoullos, Rena & Jessi Elana Aaron. 2003. Bare English-origin nouns in Spanish: Rates, constraints, and discourse functions. *Language Variation and Change* 15. 289–328. https://doi.org/10.10170S0954394503153021.

Townend, Matthew O. 2006. Contacts and conflicts: Latin, Norse and French. In Lynda Mugglestone (ed.), *The Oxford History of the English Language*, 61–85. Oxford: Oxford University Press.

Townend, Matthew O. 2002. *Language and History in Viking Age England: Linguistic Relations between Speakers of Old Norse and Old English* (Studies in the Early Middle Ages). Vol. 6. Turnhout: Brepols.

Traugott, Elizabeth Closs. 1999. A Historical Overview of Complex Predicate Types. In Laurel J. Brinton & Minoji Akimoto (eds.), *Collocational and Idiomatic Aspects of Composite Predicates in the History of English*, 239–60. Amsterdam/Philadelphia: John Benjamins Publishing Company. https://benjamins.com/catalog/slcs.47.74clo.

Treffers-Daller, Jeanine. 1994. *Mixing two languages: French–Dutch contact in comparative perspective*. Berlin/New York: De Gruyter Mouton.

Treffers-Daller, Jeanine. 2005. Evidence for insertional codemixing: Mixed compounds and French nominal groups in Brussels Dutch. *International Journal of Bilingualism* 9(3/4). 477–508.

Trips, Carola. 2014. The position proper of the adjective in Middle English: A result of language contact. In Petra Sleeman, Freek Van de Velde & Harry Perridon (eds.), *Adjectives in Germanic and Romance*, 73–94. https://doi.org/10.1075/la.212.03tri.

Trips, Carola. 2020a. Impersonal and reflexive uses of Middle English psych verbs under contact influence with Old French. *Linguistics Vanguard* 6(2). 1–14. https://doi.org/10.1515/lingvan-2019-0016.

Trips, Carola. 2020b. Copying of argument structure: A gap in borrowing scales and a new approach to model contact-induced change. In Bridget Drinka (ed.), *Historical Linguistics 2017. Selected papers from the 23rd International Conference on Historical Linguistics, San Antonio, Texas, 31 July – 4 August 2017*, 409–430. San Antonio, TX: John Benjamins Publishing Company. https://doi.org/10.1075/cilt.350.19tri.

Trips, Carola & Achim Stein. 2019. Contact-Induced Changes in the Argument Structure of Middle English Verbs on the Model of Old French. *Journal of Language Contact* 12(1). 232–267. https://doi.org/10.1163/19552629-01201008.

Trotter, David Andrew. 2003. L'Anglo-Normand: Variété insulaire ou variété isolée ? *Médiévales* 45. 43–54.

Truswell, Robert, Rhona Alcorn, James Donaldson & Joel Wallenberg. 2018. A Parsed Linguistic Atlas of Early Middle English. http://www.amc.lel.ed.ac.uk/amc-projects-hub/project/p-laeme-a-parsed-linguistic-atlas-of-early-middle-english/.

Turan, Dilek, Elena Antonova-Ünlü, Çiğdem Sağın-Şimşek & Mehmet Akkuş. 2020. Looking for contact-induced language change: Converbs in heritage Turkish. *International Journal of Bilingualism* 24(5–6). 1035–1048. https://doi.org/10.1177/1367006920926263.

Valli, André. 2007. À propos de la notion de locution verbale : Examen de quelques constructions à verbe support en moyen français: *Langue française* 156(4). 45–60. https://doi.org/10.3917/lf.156.0045.

Van Dulm, Ondene. 2009. English-Afrikaans intrasentential code switching: Testing a feature checking account. *Bilingualism: Language and Cognition* 12(2). 193–212. https://doi.org/10.1017/S1366728909004039.

Van Rooy, Bertus. 2002. Transkripsiehandleiding van die Korpus Gesproke Afrikaans.

Van Tricht, Ildiko. 2010. *De reis van Jan van Mandeville: Een onderzoek naar de relatie van de Middelnederlandse tekst met de Oudfranse traditie.* Leuven: University of Leuven MA thesis.

Velupillai, Viveka. 2012. *An introduction to linguistic typology.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Vendelin, Inga & Sharon Peperkamp. 2006. The influence of orthography on loanword adaptations. *Lingua* 116(7). 996–1007. https://doi.org/10.1016/j.lingua.2005.07.005.

Verhagen, Josje. 2009. Light verbs and the acquisition of finiteness and negation in Dutch as a second language. In Christine Dimroth & Peter Jordens (eds.), *Functional Categories in Learner Language* (Studies on Language Acquisition), vol. 37, 203–234. Berlin/New York: De Gruyter Mouton. https://doi.org/10.1515/9783110216172.203.

Verhoef, Kim, Ardi Roelofs & Dorothee J. Chwilla. 2009. Role of inhibition in language switching: Evidence from event-related brain potentials in overt picture naming. *Cognition* 110. 84–99.

Viðarsson, Heimir Freyr. 2009. *Sól gerði eigi skína:" stoðsagnir með nafnhætti í fornnorrænu.* University of Iceland MA thesis. https://skemman.is/handle/1946/2805.

Voort, Hein van der. 1995. Eskimo pidgins. In Jacques Arends, Pieter Muysken & Norval Smith (eds.), *Pidgins and creoles: an introduction*, 137–151. Amsterdam: John Benjamins Publishing Company.

Warner, Anthony R. 1993. *English Auxiliaries: Structure and History.* Cambridge: Cambridge University Press.

Weber, Andrea & Anne Cutler. 2004. Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language* 50(1). 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0.

Weinreich, Uriel. 1953. *Languages in Contact: Findings and Problems*. 7th edn. The Hague/Paris: De Gruyter Mouton.

Weinreich, Uriel, William Labov & Marvin Herzog. 1968. Empirical foundations for a theory of language change. In Wilfred P. Lehmann & Yakov Malkiel (eds.), *Directions for Historical Linguistics: A Symposium*, 95–195. Austin: University of Texas Press.

Wexler, Kenneth. 1994. Optional infinitives, head movement and the economy of derivations. In David Lightfoot & Norbert Hornstein (eds.), *Verb Movement*, 305–350. Cambridge: Cambridge University Press.

Whitney, William Dwight. 1881. On Mixture in Language. *Transactions of the American Philological Association (1869-1896)* 12. 5–26. https://doi.org/10.2307/2935666.

Wichmann, Søren. 2004. Structural patterns of verb borrowing. Conference presentation presented at the LWT workshop, Leipzig.

Wichmann, Søren & Jan Wohlgemuth. 2008. Loan verbs in a typological perspective. In Thomas Stolz, Dik Bakker & Rosa Salas Palomo (eds.), *Aspects of Language Contact: New Theoretical, Methodological and Empirical Findings with Special Focus on Romancisation Processes*, 89–121. Berlin: De Gruyter Mouton.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. 2016. New York: Springer-Verlag. https://doi.org/10.1007/978-3-319-24277-4.

Wierzbicka, Anna. 1982. Why Can You Have a Drink When You Can't *Have an Eat? *Language* 58(4). 753–799.

Wijnen, Frank. 2000. Input, intake and sequence in syntactic development. In Mieke Beers, Beppie van den Bogaerde, Gerard Bol, Jan de Jong & Carola Rooijmans (eds.), *From Sound to Sentence. Studies on First Language Acquisition*, 379–402. Groningen: Centre for Language and Cognition.

Winford, Donald. 2003. *An introduction to contact linguistics*. Oxford: Blackwell.

Winford, Donald. 2010. Contact and Borrowing. In Raymond Hickey (ed.), *The Handbook of Language Contact*, 170–187. Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781444318159.ch8.

Winford, Donald. 2013. Substrate influence and universals in the emergence of contact Englishes: re-evaluating the evidence. In Daniel Schreier & Marianne Hundt (eds.), *English as a Contact Language*, 222–241. 1st edn. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511740060.013.

Wogan-Browne, Jocelyn (ed.). 2009. General Introduction: What's in a Name: The 'French' of 'England.' In *Language and Culture in Medieval Britain: The French of England, c.1100–c.1500*, 1–14. Woodbridge: Boydell & Brewer.

Wohlgemuth, Jan. 2009. *A Typology of Verbal Borrowings*. Berlin: De Gruyter Mouton.

Wright, Laura. 1992. Macaronic writing in a London archive, 1380–1480. In Matti Rissanen, Ossi Ihalainen, Terttu Nevalainen & Irma Taavitsainen (eds.), *History of Englishes*. Berlin/Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110877007.762.

Wright, Laura. 1995. Middle English {-ende} and {-ing}: A possible route to grammaticalisation. In Jacek Fisiak (ed.), *Linguistic Change under Contact Conditions* (Trends in Linguistics. Studies and Monographs), 365–382. Reprint 2010. Berlin/Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110885170.365.

Wright, Laura. 2010. A pilot study on the singular definite articles le and la in fifteenth-century London mixed-language business writing. In Richard Ingham (ed.), *The Anglo-Norman Language and its Contexts*, 130–142. York: York Medieval Press.

Wright, Laura. 2011. Mixed-language business writing: five hundred years of code-switching. In Ernst Håkon Jahr (ed.), *Language Change: Advances in Historical Sociolinguistics*, 99–118. Berlin/New York: De Gruyter Mouton.

Zanettin, Federico. 2013. Corpus Methods for Descriptive Translation Studies. *Procedia - Social and Behavioral Sciences* 95(25). 20–32. https://doi.org/10.1016/j.sbspro.2013.10.618.

Zenner, Eline. 2013. *Cognitive Contact Linguistics: The macro, meso and micro influence of English on Dutch*. Leuven: University of Leuven PhD dissertation.

Zenner, Eline. 2017. Doomsday? (Engelse) leenwoorden in het Nederlands. In Gert De Sutter (ed.), *De vele gezichten van het Nederlands in Vlaanderen. Een inleiding tot de variatietaalkunde*, 230–248. Leuven: Acco.

Zenner, Eline, Ad Backus & Esme Winter-Froemel. 2019. Introduction: Placing usage, meaning and mind at the core of contact-induced variation and change. In Eline Zenner, Ad Backus & Esme Winter-Froemel (eds.), *Cognitive Contact Linguistics: Placing Usage, Meaning and Mind at the Core of Contact-Induced Variation and Change*, 1–20. Berlin/Boston: De Gruyter Mouton.

Zenner, Eline & Dirk Geeraerts. 2015. I'm queen of the world! (Semi-)fixed English expressions and constructions in Dutch. *Taal & Tongval* 67(2). 247–274.

Zenner, Eline, Kris Heylen & Freek Van de Velde. 2018. Most borrowable construction ever! A large-scale approach to contact-induced pragmatic change. *Journal of Pragmatics* 133. 134–149.

Zenner, Eline, Dirk Speelman & Dirk Geeraerts. 2012. Cognitive Sociolinguistics meets loanword research: Measuring variation in the success of anglicisms in Dutch. *Cognitive Linguistics* 23(4). 749–792. https://doi.org/10.1515/cog-2012-0023.

Zenner, Eline, Dirk Speelman & Dirk Geeraerts. 2015. A sociolinguistic analysis of borrowing in weak contact situations: English loanwords and phrases in expressive utterances in a Dutch reality TV show. *International Journal of Bilingualism* 19(3). 333–346. https://doi.org/10.1177/1367006914521699.

Zenner, Eline & Dorien Van De Mieroop. 2017. The social and pragmatic function of English in weak contact situations: Ingroup and outgroup marking in the Dutch reality TV show Expeditie Robinson. *Journal of Pragmatics* 113. 77–88. https://doi.org/10.1016/j.pragma.2017.02.013.

Zimmerman, Richard. 2018. The Parsed Corpus of Middle English Poetry.