Academic Year
2021 - 2022

The *go*-futures in English and French: a corpus-based diachronic contrastive analysis

On the influence of translations on the conventionalization of the future semantics in [BE *going to* INF] in Early Modern English

Pauline Claes

Master of Linguistics (Digital Text Analysis)

Master Thesis

Promotor
Prof. dr. Peter Petré

Assessor
Prof. dr. Astrid De Wit

Universiteit Antwerpen
Faculteit Letteren
en Wijsbegeerte

Table of contents

List of Tables

List of Figures

The *go*-futures in English and French: a corpus-based diachronic contrastive analysis

On the influence of translations on the conventionalization of the future semantics in [BE *going to* INF] in Early Modern English

Pauline Claes
Master of Linguistics (Digital Text Analysis) 2020–2021 (2022)
University of Antwerp
Faculty of Arts
pauline.claes@student.uantwerpen.be
Student ID: 20163274

Prof. dr. Peter Petré
Promotor
University of Antwerp
Faculty of Arts
Mind-Bending Grammars
peter.petré@uantwerpen.be

ABSTRACT. *This paper studies the conventionalization processes of the go-futures in French and in English contrastively, and the influence that the translation practice of French into English might have had on the conventionalization of the future semantics in [*BE *going to INF] primarily, and on the use of [*GO *to V] secondarily in Early Modern English in light of the contact hypothesis. To this extent, the use of progressive and non-progressive go-constructions as well as the contexts in which these occur were analyzed in a corpus of 10 million words consisting of translations, their French source texts, and non-translated English texts from two time frames (1580-1609 and 1680-1699). This research points toward an observable influence of French source texts on the use of progressive and non-progressive go-constructions in translations, while the influence on non-translated language remains ambiguous.*

KEYWORDS: *conventionalization, future semantics, contact hypothesis, corpus-based analysis.*

Plagiarism statement. The undersigned, Pauline Claes, student of the Master of Linguistics (Digital Text Analysis) at the University of Antwerp, declares that this thesis is completely original and exclusively written by the undersigned. For all information and ideas derived from other sources, the undersigned has referred to the original sources, both explicitly and in detail.

Antwerp, 10 januari 2022

Number of words: 18,893

1. Introduction

The current paper investigates the potential influence of the French *go*-future [ALLER INF] (*futur proche*) on the conventionalization of the future semantics in the English [BE *going to* INF] construction primarily and on the use of the non-progressive [GO to V] secondarily, through translations from French into English across two time frames in Early Modern English: EARLY (1580-1609) and LATER (1680-1699). While this has already been studied extensively in English, the relation to the development of the French *go*-future has not yet received much attention. Given that both *go*-futures arose from lexical motion verbs and developed into a motionless future marker along similar patterns, and considering that the development of the French *go*-future chronologically preceded that of the English one, a contrastive study might provide new insights. This paper aims to fill that need. To this extent, a corpus of 10 million words consisting of translations, their corresponding French source texts, and non-translated reference texts was collected, annotated, and analyzed. More specifically, the frequency of use of the constructions themselves as well as of the contexts in which these occur were compared between text types (translation, reference, source texts) and between time frames. The development of lexical motion verbs into grammatical future markers is a common process (Pérez 1990; Bybee, Perkins, and Pagliuca 1994), and contact between languages can only be assumed to be the source of a linguistic change when that change would be unlikely to have occurred without the contact situation (Thomason 2003), so one must be careful in drawing any solid conclusions with regard to the contact hypothesis (Danchev and Kytö 1994; Danchev and Kytö 2002). Nevertheless, this research points toward an observable influence of French source texts on the use of the progressive and of the non-progressive *go*-construction in translations, as well as on the contexts in which these occur. Still, the direct influence of French source texts and of translations on non-translated language remains an open question. While it is unlikely that French has been the source of the grammaticalization process in English, the existence and presence of *futur proche* in source texts arguably might have accelerated the development of [BE *going to* INF] to a future marker in English through translations, while the foundations for this development were already in place.

First, in section 2, I will provide an overview of the literature on grammaticalization and constructionalization in general, of the history of [BE *going to* INF] and [ALLER INF] in specific, and how their developments have followed similar patterns. This will lead me to the study of contact-induced grammaticalization and the contact-hypothesis. Based on this theoretical framework, I will present the hypotheses that I have studied in 2.5. Section 3 will elaborate on the corpus that has been collected and the methods used to make it operationalizable. The results of the analyses will be described and discussed in section 4. The challenges faced in this research will be discussed in 5. Lastly, section 6 sums up the main findings and section 7 points toward avenues for future research.

2. Theoretical framework

In this section, I will provide a theoretical framework for studying the *go*-futures in English and in French. First, I will define grammaticalization and constructionalization in 2.1, and the mechanisms of change which these studies have typically identified, as well as how these translate to more recent literature on constructionalization in 2.2. After that, the conventionalization of *go*-futures in general will be discussed (2.3), and of [BE *going to* INF] (2.3.1) and [ALLER INF] (2.3.2) in specific. This will lead me to the

contact hypothesis and contact-induced language change, which will be discussed in 2.4. Lastly, in 2.5, I will discuss the hypotheses that will be investigated in this research.

2.1 Definition

Traditionally, 'grammaticalization', also known as 'grammaticization' or 'grammatization', has been defined as the study of the sources of grammatical forms and the stages of change which they undergo subsequently (Hopper and Traugott 2003). The term was first defined by Antoine Meillet as 'the attribution of grammatical character to an erstwhile autonomous word' (Meillet, 1912, as cited in Hopper and Traugott, 2003). This definition is, however, too narrow, as the process of grammaticalization is not confined to the development of lexical forms into grammatical forms, since grammatical forms can also become more grammatical (Haspelmath 1999; Heine 2003). Linguists have been studying language in general, and grammaticalization in particular, from two main perspectives: the synchronic approach and the diachronic (or historical) approach. While the former studies language at a single point in time, the latter does so between two or more points in time. The synchronic perspective views grammaticalization as primarily a syntactic, discourse pragmatic phenomenon studied from the point of view of fluid patterns of language use (Hopper and Traugott 2003). The diachronic perspective is the most prominent in grammaticalization studies, and Hopper and Traugott (2003) define grammaticalization in this perspective as 'that subset of linguistic changes whereby a lexical item or construction in certain uses takes on grammatical characteristics, or through which a grammatical item becomes more grammatical' (2003, 2). Heine (2003) underlines that, despite there existing both synchronic and diachronic dimensions within the research field, the study of grammaticalization should be considered to be essentially diachronic in nature.

In more recent literature, however, scholars subscribing to a constructionalist view on grammar propose a different approach to studying language change, namely 'constructionalization'. While sharing some common ground, 'grammaticalization' and 'constructionalization' are conceptually different to some extent (Trousdale 2014). Traugott and Trousdale (2013) define constructionalization 'as the creation of a form$_{new}$-meaning$_{new}$ pairing, in other words, as the development of a new sign' (2013, 1; 22). The model of constructionalization which they propose aims to identify from a construction grammar perspective how new constructions arise and assume different functions. They distinguish two types of constructionalization: lexical (or contentful) constructionalization and grammatical (or procedural) constructionalization. In their view, lexical grammaticalization involves constructional change in which a new semantically non-compositional contentful form arises from a productive compounding pattern in the grammar, which is new in both semantics and morphosyntax. Grammatical constructionalization is defined as the changes in form and meaning which create constructions that encode grammatical form$_{new}$-meaning$_{new}$ pairs that differ from their lexical sources in being less referential, and more grammatical than their sources (Traugott and Trousdale 2013, 22–26). It is important to understand that these two seemingly opposite types of constructionalization are actually located at the 'poles of the contentful-procedural gradient' (2013, 22), and that many constructions will therefore take an intermediate position on this continuum, possibly displaying both properties associated with lexical constructionalization as well as properties associated with grammatical constructionalization (Trousdale 2014, 562–563). In this continuum, and thus in both types discussed here, gradualness is a key concept. This means that constructionalization is characterized by a succession of conventionalized incremental

micro-steps that can precede and follow constructionalization (2013, 26). Traugott and Trousdale (2013) name these steps 'constructional changes', and define them as changes 'affecting one internal dimension of a construction' (2013, 26). Next to the gradualness of constructionalization, these constructional changes are also an important notion in the theory on constructionalization as presented by Traugott and Trousdale (2013), because they emphasize that one must be able to identify the conventionalization of changes in both form and in meaning to be able to speak of 'constructionalization'. Otherwise, we should speak of a 'constructional change' (2013, 22–26). They illustrate this difference by the idea that innovations are brought about by individuals, but need to be supported by and shared across individual networks in a population, largely through processes of neoanalysis and analogization (2013, 46). Neoanalysis and analogization will be further discussed in 2.2.1. Therefore, they argue that most change actually consists of conventionalized constructional changes, as they only affect either the form or the meaning of an existing construction (node), and not both (2013, 91).

In this light, Petré (2019) notes that, in the case of the development of the lexical source string 'BE *going to* INF' to the construction [BE *going to* INF] as a deictic future marking auxiliary, there does not seem to be a noticeable formal difference, at least not in the early instances (the data analyzed in Petré (2019) extends to 1700). However, there is a definitive semantic difference to be observed from the lexical source expressing 'controlled motion with a purpose', in that there is an extension to inanimate subjects and motionless contexts. In this regard, Petré (2019) disagrees with the view on constructionalization as proposed by Traugott and Trousdale (2013), arguing that the independent developments might have been connected and might have led to a new global cognitive schema for [BE *going to* INF], and that therefore, in this view, constructionalization essentially does not require a change in both form and in meaning, but a semantic shift alone may suffice (2019, 27).

In conclusion, the notions of 'grammaticalization' and 'constructionalization' both investigate the extent to which (constructional) language changes conventionalize, albeit from a slightly different theoretical angle. Going forward in this paper, I will use 'conventionalization' in order to avoid any ambiguity. Given that many of the mechanisms of change identified in the process of conventionalization have been defined in research on grammaticalization, the latter term will resurface when discussing these.

2.2 Mechanisms of change

In this section, I will discuss the main mechanisms of change that have traditionally been discussed in grammaticalization studies, and how these translate to studies on constructionalization. I will do so in a hierarchical manner, going from more general to more specific. First, I will treat neoanalysis (reanalysis) and analogization (analogy) in 2.2.1. After that, (uni)directionality, the notion of extravagance and the role of frequency will be discussed alongside each other in 2.2.2. Subsequently, I will discuss the semantic mechanisms of change of reduction and expansion of meaning and how these fit into the loss-and-gain model in 2.2.3.

2.2.1 Neoanalysis (reanalysis) and analogization (analogy). Neoanalysis ('reanalysis') and analogization ('analogy') are two general mechanisms at play in the process of grammaticalization. However, their relation to each other and their respective roles in grammaticalization remain rather controversial.

'Reanalysis' was defined by Langacker (1977, 58) as 'change in the structure of an expression or class of expressions that does not involve any immediate or intrin-

sic motivation of its surface manifestation', and was long considered to be the most prominent mechanism in the process of grammaticalization. 'Analogy' is said to be an overt process that brings the change, brought into effect by reanalysis, to the attention of the language users and linguists, thus making the unobservable changes of reanalysis observable (Hopper and Traugott 2003, 68). However, some scholars have disagreed and argue in this respect that it is analogy that should be regarded as the prominent mechanism at play in grammaticalization (Anttila 2003; Fischer 2007; De Smet 2009, as cited in Traugott and Trousdale 2013, 37–38).

Traugott and Trousdale (2013) treat these two concepts and their relation more carefully. First of all, they assure that 'neoanalysis', as proposed by Andersen (2001, as cited in Traugott and Trousdale 2013), is a more accurate term for this mechanism. Evidently, a language user cannot 're'-analyze a construction that they had not yet internalized before, but rather come to a 'different' analysis (2013, 36). Secondly, they suggest that the mechanism previously called 'analogy' be changed to 'analogization', in order to avoid any ambiguity between the mechanism of analogy and the process of analogical thinking, which they identify as a motivation of change rather than a mechanism of change. They define analogization as 'a mechanism or process of change bringing about matches of meaning and form that did not exist before' (2013, 38). Lastly, they conclude that constructions always, despite some exceptions, have a link to an already existing one, rather than being entirely new. In that regard, they question the view that analogy is the primary mechanism of change. Instead, they argue that analogization is neoanalysis because it necessarily entails micro-step changes. Therefore, they view neoanalysis as the primary mechanism of change in conventionalization (2013, 58).

2.2.2 General mechanisms of change: (uni)directionality, the notion of extravagance, and the role of frequency. Typically, the process of grammaticalization is assumed to be unidirectional. This concretely means that the process of grammaticalization leads a morpheme or construction from 'less' to 'more grammatical', and not vice versa (Heine, Claudi, and Hünnemeyer 1991; Traugott and Heine 1991). In that respect, the notion of unidirectionality has been argued to be systematic and cross-linguistically replicated (Hopper and Traugott 2003), and to be one of the most important constraints on language change in general, and on possible syntactic changes in particular (Haspelmath 1999), of which counterexamples remain rare and are often refuted (Heine 2003).

However, more recent studies treat this traditional view on (uni)directionality in a more critical manner. Traugott and Trousdale (2013) argue that this directionality is a mere observation and is not inherent in grammaticalization (2013, 100). Accordingly, even though the directionality from a procedural (grammatical) to a contentful (lexical) function is indeed less frequent than the directionality from a contentful (lexical) to a procedural (grammatical) function (2013, 124), they argue that Haspelmath (1999)'s presupposition that grammaticalization only involves change from lexical to grammatical status should be questioned (2013, 126).

The role of directionality in conventionalization should be discussed alongside the notion of 'extravagance', a term coined by Haspelmath (1999). Both speakers and listeners play an important role in communicative situations by manipulating and interpreting meaning. While the listener is actively interpreting and disambiguating speech, the speaker's main goal is to communicate: being heard and interpreted. Therefore, the speaker will seek ways to attract attention and be socially successful (Traugott and Trousdale 2013, 124), and to guide the listener in their interpretation, thus seeking to enhance expressivity through new and innovative ways of conveying the same message (Hopper and Traugott 2003). When language users have the desire to be noticed,

and want their statement to stand out among other statements, they might want to make it more emphatic, through intonation, unconventional and unexpected language. Accordingly, this tendency makes individual language users ultimately responsible for syntactic change, and the notion of extravagance in this sense constitutes an important motivation for this (Petré 2016b, 115–116; 2017, 2).

It is important to understand that frequency arguably plays an important role in this as well. When individual language users have the urge to stand out, to achieve social success, they use unexpected, 'extravagant' language and bring about innovations in that way (Haspelmath 1999, 1057; Petré 2016b, 118). However, if they succeed, the innovative language they used might get imitated and adopted by the linguistic community. This spread then logically might result in an increase in frequency, possibly causing routinization of this unconventional form, thus rendering it conventional. Evidently, this calls for a new search for an extravagant or unconventional expression. This is known as 'renewal' (Hopper and Traugott 2003, 122, as cited in Petré 2016b, 118). Importantly, Petré (2016b) notes that it is arbitrary and unpredictable when in history renewal might occur. This phenomenon is what Petré (2016b) defines as a 'cycle of convention and innovation'. However, it must be understood that this going back and forth between convention and innovation does not mean that a form that has become conventional regains its unconventional status. Instead, it means that the search for a new innovative form has reopened. In fact, it is highly unlikely that a form regains its unconventionality after having become a convention. Accordingly, this imbalanced proportion may explain the unidirectionality typically assumed in grammaticalization (Petré 2016b, 142).

Additionally, this interaction between convention and innovation, and the role which frequency arguably plays in it, fits into an important discussion on the role of frequency in the literature on grammaticalization studies. Frequency has been argued to play an important role in the process of grammaticalization, in that the grammaticalizing morpheme or construction typically (though not necessarily always) demonstrates a dramatic increase in text frequency. In this perspective, Bybee (2003) has argued that this increase in frequency is not only a direct result of the process of grammaticalization, but is also a primary contributor to it, since an increasing frequency might result in the linguistic item losing in semantic meaning, undergoing phonetic reduction, and being used in more contexts, which in turn possibly results in an even higher frequency (2003, 604, 621–622). These innovations typically characterize the later stages of grammaticalization (Traugott and König 1991), and it must be noted that an increasing or high frequency can also result in linguistic items being highly entrenched, and therefore less likely to change (2003, 619). Yet, some scholars discuss the role of frequency in a more critical manner. For example, Hoffmann (2004) argues that a linear connection between the level of entrenchment and the frequency of occurrence is an oversimplification. Consequently, while the frequency of a linguistic item might be meaningful as a diagnostic tool when compared with the frequency of occurrence of related linguistic phenomena, it should not be attributed an absolute status. It might therefore seem useful to use a more differentiated approach to frequency-based analysis, such as the one Hoffmann proposes, in which it should not only be determined how often a particular item is actually found, but also how often it could have occurred but was expressed differently. Yet, as concepts can often be realized and expressed in many different and even formally unrelated ways, it is evident that this is rather difficult to retrieve in an automated way (2004, 191–195).

Traugott and Trousdale (2013) subscribe to Hoffmann's critical view on the matter, and supplement it in arguing that 'sufficient' frequency is not operationalizable (2013,

11), because it is simply problematic to determine the level of frequency needed for pattern storage and entrenchment, especially in historical studies where the textual material at hand is minimal (2013, 5).

However, this point of view is challenged by the method that Petré (2019) proposes to operationalize frequency. They argue that, in the case of 'BE *going to* INF', the patterns discerned from the run-up phase towards constructionalization systematically background or deprofile certain lexical aspects. Essentially, where the string was previously used with animate, agentive subjects in a context of motion, the gradual conventionalization allows for the construction to also be used in motionless contexts with inanimate, non-agentive subjects (2019, 3). These patterns are defined as 'assemblies': 'recurrent configurations of existing constructions and their co-text/context, which do not yet have constructional status themselves' (2019, 3). Consequently, a quantitative analysis of the contexts in which these assemblies occur provides a better measure for the emergence of a construction, as it is the frequency shifts in those assemblies that may lead to change (2019, 3–4).

2.2.3 Semantic mechanisms of change: reduction and expansion of meaning and the loss-and-gain model. Typically, in grammaticalization studies, it is argued that the process of grammaticalization results in a loss in specific features of meaning (Bybee 2003; Heine 2003). This is commonly referred to as 'bleaching', 'desemanticization', and 'semantic reduction'. It means that concrete meanings become reinterpreted as more abstract grammatical meanings in specific contexts, and thus become more general and more abstract in their meaning, allowing them to be more widely applicable and more frequently used (Bybee 2003). This mechanism comes down to the idea that the force or impact of a stimulus weakens if it is used very frequently, resulting in generalization of meaning (Haiman 1994; Bybee, 133).

The concept of 'bleaching' is, however, not uncontested. Some scholars rightly point out that this would inappropriately suggest an impoverishment, as grammatical meaning is added (Sweetser 1988; Traugott and König 1991). On that note, Sweetser gives the important example of the grammaticalization of the verb *go*, in which the sense of physical motion is lost, while the new meaning of future prediction or intention is added (1988, 392). This is known as the 'loss-and-gain' model (Traugott and Trousdale 2013, 105). It concretely means that in the process of grammaticalization, we typically see a generalization of the lexical meaning of the source domain (*loss*), but an addition of grammatical meaning to the target domain (*gain*) (Sweetser 1988).

Furthermore, in this respect, Traugott and Trousdale (2013) discuss the distinction between 'grammaticalization as reduction and increased dependency (GR)', and 'grammaticalization as expansion (GE)'. The view on grammaticalization as reduction and increased dependency is usually associated with earlier work on grammaticalization, and refers to the reduction of the structural complexity and semantic substance of a grammaticalizing item, as well as the increase in bondedness on a paradigmatic and syntagmatic level (2013, 102). Scholars subscribing to the GE perspective (such as Bybee, Perkins, and Pagliuca 1994, as cited in Traugott and Trousdale 2013, 106) argue that the generalization of meaning ('bleaching') of a grammaticalizing linguistic item is associated with a wider use and an expanded variety of contexts. While not stated explicitly, Traugott and Trousdale (2013) rather seem to subscribe to the 'loss-and-gain' model, and argue that, indeed, some lexical meaning is lost but grammatical meaning is added. In their view, these two perspectives might seem irreconcilable at first, but they are actually largely complementary, since both are focused on a different aspect of grammaticalization. While GR mostly answers questions about the development of

morphosyntactic form, GE is focused on changes in a linguistic item, as well as on how grammaticalization occurs in contexts (2013, 109).

In sum, a 'loss-and-gain' model seems more appropriate when studying the conventionalization of grammatical forms, as the reduction of lexical meaning is essentially associated with an increase in grammatical meaning. A *reduction* of lexical specificity allows a linguistic item to be used in more contexts, resulting in an *expanded* use.

2.3 The conventionalization of *go*-futures

Bybee, Perkins, and Pagliuca (1994) note that there is a restricted range of lexical sources from which futures can evolve. More specifically, cross-linguistically speaking, futures arise generally from constructions involving movement verbs, from markers of obligation, desire, and ability, and from temporal verbs. In fact, movement verbs constitute the most prominent source for futures than other verbs or lexical material of any kind. 'Come' and 'go' are examples of such movement verbs. In the majority of cases, the movement verb in question is expressed as an auxiliary, outnumbering the cases where the movement verb is expressed as a suffix, a prefix or a particle (Bybee, Perkins, and Pagliuca 1994).

They also highlight that simple movement without an allative component of 'movement toward' does not generally evolve into future marking. Generally, it is assumed that the temporal meaning is inferred from the spatial meaning: when one moves in space, one also moves in time (Disney 2009a, 64). The major change occurs however when the spatial meaning gets lost, and the expression of intention becomes more prominent. As, from then on, the construction expressing future can be extended to contexts in which an intention is expressed without movement (Bybee, Perkins, and Pagliuca 1994). In that respect, De Mulder (2008, 360–361) has noted that, the notion of the destination moves more to the background once the notion of the intention becomes more important in the context, to such an extent that the destination may even become redundant. Subsequently, the notion of intention is backgrounded once the meaning of the construction is generalized and extended to contexts allowing for inanimate subjects incapable of movement. This is in line with Bybee and Pagliuca (1987), who argue that movement verbs initially require a subject capable of movement. This original sense, however, weakens gradually until the point where the marker signals prediction, and becomes applicable in contexts with any sort of subject. Yet, as argued in Bybee (2003), this original sense is not entirely lost, but retained in some contexts. This arguably results in futures that derived from different sources will convey different meanings (Bybee and Pagliuca 1987).

In fact, Pérez (1990) notes that it is a 'universal channel of grammaticalization' for a construction to develop from movement > intention > prediction. Or, 'where *be going to* initially signifies that the subject is starting out from a present position in space towards a physical goal, it acquires, through the notion of intention, the meaning of starting at a present point in time and travelling towards a future goal' (1990, 4).

In what follows, I will provide an outline of the respective histories of the conventionalization of future meaning in [BE *going to* INF] in English and [ALLER INF] in French. It is clear that these two constructions demonstrate similar patterns and similar developments, even though the French construction chronologically preceded the English one. Given that there was an important translation practice of French texts into English (Danchev and Kytö 1994), there is reason to believe that there might have been an influence from the French construction to some extent. However, this will be discussed in more detail in 2.4.

2.3.1 The conventionalization of the future semantics in 'BE *going to* INF' in English. In this section, I will provide an outline of the history of the grammaticalization of 'BE *going to* INF' as a future marker in English, which has been studied extensively, especially in grammaticalization or constructionalization studies. It evolved from expressing 'movement' to 'intented purpose' to 'non-movement intention' to 'prediction', to the point where it contrasts with *will* in Present Day English (Disney 2009a, 63–70). Due to the limited nature of this paper, I cannot go into much detail here. For a detailed account of the history and development of the *be going to* construction, please refer to Disney (2009a,b), Budts and Petré (2016), Petré and Van de Velde (2018), Petré (2019).

The composing elements of the construction originated in Old English. The verb *go* is both traced to *gān* as well as *gangan*. Although the origin and relation between these two is not entirely clear, they both convey a general motional meaning (Pérez 1990; Budts 2014). As already discussed, a more general meaning makes a linguistic item more easily extended into other semantic domains. While there are other verbs expressing movement as well, such as *walk, run,* or *wander*, these all imply which type of movement is involved, and consequently which type of subject that is able to perform the action. By contrast, *go* remains agnostic about the manner of movement, and this is arguably the main reason why *go* appears in this construction (Pérez 1990). In addition, the progressive reinforces an interpretation of ongoingness. While the origins of the English progressive constructions are ambiguous as well, it is argued that the [BE V*ing*]-construction expressing an ongoing action occurred around 1600 (Pérez 1990; Budts 2014; Petré 2016a, 2017). For a more detailed account on why [BE V*ing*] became the progressive and the grammaticalization of the progressive, please refer to Elsness (1994), Nesselhauf (2007), Petré (2016a, 2017), and De Wit, Petré, and Brisard (2020).

The conventionalization of future meaning in [BE *going to* INF] has received much attention in grammaticalization studies. It developed from a compositional string expressing motion with a purpose to an auxiliary expressing future, with the purposive adjunct as its complement (Petré and Van de Velde 2018, 875). However, the chronology of the developments, especially the early ones, is less straightforward.

Danchev and Kytö (1994) mention that most authors regard the example given in *Oxford Early Dictionary*, found in 1482, to be the first instance of the construction:

1. "Therfore while thys onhappy sowle by the vyctoryse pompys of her enmyes *was goyng to be broughte* into helle for the synne and onleful lustys of her body. Loe sondenly anon came done an hye fro heuyn a gret lyght" (The Revelation to the Monk of Evesham 1482: 43, Arber, 1869, as cited in Danchev and Kytö, 1994)

However, Petré and Van de Velde (2018, 875) note that this attestation is often wrongfully interpreted as expressing motionless future. It is because of the passive that the notion of agency or control is weakened, and it is because of the use of the goal in the infinitival clause, that the directionality or motion is weakened. Therefore, it should not be considered the first attestation of the semantic redistribution of 'BE *going to* INF' (Hopper and Traugott 1993, 2003; as cited in Petré and Van de Velde 2018, 875). It could, however, serve as an indication that the opportunity of semantic redistribution is there (Petré and Van de Velde 2018, 875), or that some pragmatic ambiguity arose between motion with a purpose and futurity, possibly invoking a use in motionless contexts later on (Traugott and Trousdale 2013, 219).

A more interesting instance (2) is one made by grammarian Poole in the first half of the seventeenth century. This attestation demonstrates not only that *going to* could be

used as an equivalent of *about to*, indicating that the former was already appropriate in motionless contexts at that time, but also that it must have already been conventionalized to a certain extent if a grammar book for general use included it (Petré and Van de Velde 2018, 875).

2. About to, or *going to*, is the signe of the Participle of the future...: as, my father when he was about [to] die, gave me this counsell. I am [about] or *going [to]* read.
(Poole 1646, 26; as cited in Danchev and Kytö 1994, 67; square brackets in the source text)

Above, I argued that (1) cannot be the first attestation of the use of [BE *going to* INF] expressing motionless future. However, this instance is noteworthy for another reason. Since it is an English translation of Latin (Danchev and Kytö 1994), it raises the question of the relationship between source texts and translations: was the translation influenced by the paraphrasing of the original, or did the translator make use of a form already known in the target language of that time? This is a difficult question to answer. Nevertheless, Pertejo (1999) remarks that most of the earlier examples of the construction are found in translations from other languages, mainly Latin and French. Another early instance is (3), out of an English translation of a French text, cited by Mossé (1938, as cited in Danchev and Kytö, 1994). Danchev and Kytö (1994) provided the corresponding sentence from the French source text (4).

3. "... 'sir', quod Gerames, 'we be frenchmen, pylgrymes, & *are goyng to offre* at y^{e} holy sepulcre. ..."
(1534, *Huon of Burdeux*, 191; as cited in Danchev and Kytö, 1994, 62)

4. Sire ce dist gerasmes nous sommes francois qui *venons de aourer* le sainct sepulchre
(1516, *Huon of Burdeux*: xlviii recto; as cited in Danchev and Kytö, 1994, 62)

Danchev and Kytö (1994) rightly point toward the potential influence of the movement verb present in the French source text on the translation in English, in that the use of a motion verb in the source text might have determined the use of a motion verb in the translation. However, it must be noted that the use of the motion verb in the French source text here is not an instance of *futur proche* (ALLER + INF; expressing (near) future), but rather an instance of *passé récent* (VENIR DE + INF; expressing recent past). While both are indeed movement verbs, they express different tenses. Therefore, it might be better to be careful in drawing any conclusions in that respect.

A similar implication is raised in Petré and Van de Velde (2018, 876). They mention a passage from Henry Ainsworth's commentary on the Old Testament, dated 1616, containing '*going to dye*'. Traugott and Trousdale (2013) interpreted this occurrence as a marker of futurity rather than a motion verb. Petré and Van de Velde remark that, at first glance, the instance indeed suggests a motionless future, but that the idea of motion is actually inherently strong in the Hebrew verb that was translated with 'going'. Therefore, they hypothesize that Ainsworth, as a knowledgeable Hebrew scholar, might have wanted to be as faithful as possible to the Hebrew source text, and that this example should not be seen as an indication of the conventionalization of the future meaning in 'going'. In fact, given that Ainsworth deemed it necessary to explain this

use in a note, this points toward the conclusion that it was not conventional in English at the time (2018, 876).

In sum, even though the first attestations of the construction might remain contentious, Petré and Van de Velde conclude that the conventionalization of [BE *going to* INF] most likely occurred between 1620 and 1640, reasoning that a motionless *'be going to* INF' only conventionalized some years after Ainsworth's translation, and assuming that a general grammar would probably not include a brand-new construction (2018, 877). Furthermore, English translations from source texts in other languages raise questions about the influence of the source language as well as the influence of the translator or their style. This leads me to the contact hypothesis, which will be discussed alongside theory on contact-induced language change in 2.4.

2.3.2 The conventionalization of the future semantics in 'ALLER INF' in French. The French *go*-future consists of a conjugation of the verb 'aller' (*go*) followed by an infinitive and is commonly called *futur proche* or *futur périphrastique*. The construction and its conventionalization process have been researched in the past, although not as extensively and thoroughly as the English [BE *going to* INF] construction. The construction of [ALLER INF] has grammaticalized to the extent that it is recognized as a full tense in contemporary grammars. In fact, it is the most transparent periphrastic construction in French, in that it does not impose any restrictions whatsoever on the infinitive that follows (Vetters and Lière 2009, 32).

It developed from a lexical verb, *aller*, denoting movement in space towards a destination. According to De Mulder (2008), the development of the lexical motion verb to a grammatical future marker took place in two metonymic shifts. In the first metonymic shift, between the thirteenth century and the first half of the sixteenth century, the construction developed a notion of intention. This was first expressed through the act that the agent of the motion will carry out when arriving at the destination, as becomes clear from example (5). However, when the intention is actually more important in the context than the destination where it will be carried out or when the destination can be inferred from the context, the need for the specification of the destination becomes redundant, as in example (6). Interestingly, these changes particularly occur in contexts with a subject in the first person singular (De Mulder 2008, 360–361).

5. Car incontinant le roy manda tous ses barons, cappitaines et chefz de guerre, et sans aulcun delay fit appareillier tout ce qui estoit de besoing pour *aller* en Espaigne *commancer la guerre contre les barons du pays*.
(Jehan de Paris 8, as cited in De Mulder 2008, 360)
[For immediately the king mandated all his barons, captains and warlords, and without any delay made all that was needed *to go* to Spain *to start the war against the barons of the country*.]

6. Il est bien temps de deviser / Les personnaiges et nommer. / Je vous les *veulx* nommer à tous. / Je *voys* au Monde commencer.
(*Moralité de Charité*, 1532–1550, as cited by Gougenheim 1929, 98; De Mulder 2008, 360)
[It's time to discuss / The characters and to name / I want to name them all to you / I'*m going* to the World to start.]

Yet, at this stage, the construction of 'ALLER INF' is not yet a grammatical future marker. This development only starts in the second metonymic shift, taking place in

the first half of the sixteenth century, which is based on the idea that the realization of intentions generally happens in the future (De Mulder 2008, 361). Importantly, this second shift brings about a generalization of meaning, allowing the verb *aller* to be combined with infinitives that no longer depend on the subject's intention, resulting in [ALLER INF] only expressing future (7) (De Mulder 2008, 361–362).

7. Par deffaulte de patience, / Tu *vas perdre ta conscience*.
(*Moralité de Charité*, 1532–1550, as cited in Gougenheim 1929, 98; De Mulder 2008, 362)
[For lack of patience, you *are going to lose* your conscience.]

At this stage, [ALLER INF] is arguably still distinguishable from *futur simple* (De Mulder 2008, 362), in that only the former is able to express relative future from the perspective of the speaker (Vetters and Lière 2009, 28). Nevertheless, we can conclude from (8) that its use with a future meaning has found its way to literary and formal discourse in the course of the sixteenth century, even to the extent that it was used instead of *futur simple* in some contexts. To be more precise, in (8), [ALLER INF] is used to announce a story, whereas other authors at that time would have used *futur simple* or *vouloir* (Gougenheim 1929, 99). The source text of (8) is included in this research.

8. "Pour ce, mes Dames, que je me suis souvent souzhaicté Compaignon de la fortune de celuy dont je *vois* faire le compte, je vous diray que en la ville de Naples..."
(1559, Marguérite de Navarre, *L'Heptaméron*, 273, as cited in Gougenheim, 1929, 99)
[For this, Ladies, that I have often wished myself Companion of the fortune of the one of which *I am going to make* the account, I will tell you that in the city of Naples...]

In addition, it is argued that the future sense is secured only in the beginning of the seventeenth century, when combinations with inanimate subjects become appropriate (9, 10) (Bres and Labeau 2018, 57).

9. La paix *va refleurir*, les beaux jours *vont renaître*.
(Racine, *Andromaque*, II, 4, 1667; as cited in De Mulder 2008, 362)
[Peace *is going to bloom* again, beautiful days *are going to be reborn.*]
(*(Re)naître*, being born, is an active verb in French.)

10. Apres tant de travaux, apres tant de combats, / Rome *s'en va tomber* sous l'effort de ton bras. [...]
(G. de Scudéry, *Alaric, ou Rome vaincue*, Paris, A. Courbé, 1654, p.375; as cited in Bres and Labeau 2018, 57).
[After so much work, after so many battles, Rome *is going to fall* under the effort of your power.]

However, a brief qualitative analysis of the source text corpus included in this research points toward a different conclusion. To be more precise, (11) and (12) are clear examples retrieved from French texts in the source text corpus of this research, where the subject is inanimate, and no motion or intentionality are expressed:

11. Et en disant hin, la bonne femme, de l' aise qu' elle avoit en son compte, se print à faire la ruade que feroit son poulain : et en la faisant *sa potée de lait va tomber*, et se respandit toute : Et voila ses eufz, ses poussins, ses chapons, ses cochons, sa jument et son poulain tous par terre.
(Des Périers, Bonaventure, *Nouvelles récréations et joyeux devis*, 1558)
[And being surprised, the good woman, by the ease that she had in her expenses, suddenly rears up like her foal would: and in doing so, *her pot of milk is going to fall*, and spilled all over: And lo and behold her eggs, her chicks, her capons, her pigs, her mare and her foal, all on the ground.]

12. Nous tenós auec Aristote, Galien & la verité mesines, que la reception se fait au crystallin, pource que c'est la plus noble partie de l'œil, ayant vne substance toute particuliere, estant situé au milieu de l'organe comme au centre, ou *se vont rencontrer les deux lumieres*, l'exterieure, qui entre par la prunelle comme par vne fenestre, & l'interieure qui est apportee par le nerf optique.
(Du Laurens, André, *Discours de la conservation de la veüe: des maladies melancholiques: des catarrhes: et de la vieilless*, 1597)
[We hold with Aristotle, Galen and the [*"mesines"*: unknown] truth, that the reception is made in the crystalline, for it is the noblest part of the eye, containing a very particular substance, being located in the middle of the organ as in the center, where *the two lights are going to meet*; the external one, which enters the prunella as it would a window, and the interior one which is brought by the optic nerve.]

This indicates that the future sense of *futur proche* with inanimate subjects was already in place at least a century earlier than is assumed in De Mulder (2008); Bres and Labeau (2018). This will be analyzed and discussed quantitatively in the following sections.

Instance (10) is also of particular interest, because it demonstrates that *s'en aller* is used interchangeably (although less frequently) with *aller* in this construction at this time (Gougenheim 1929; Bres and Labeau 2018, 58). Since it was dismissed in the eighteenth century (Bres and Labeau 2018, 58), this is no longer the case in present-day French, but it is still used in the period of the data studied in this research, as can be seen in (12).

In conclusion, 'ALLER INF' underwent similar developments as 'BE *going to* INF', in that it developed from a lexical motion verb to a grammatical future marker. Furthermore, the future semantics in 'ALLER INF' conventionalized at least a century earlier than the English construction. Considering that a large number of texts were translated from French into English during the Later Middle English and Early Modern English periods (Danchev and Kytö 1994), there might have been a substantial influence of the French *futur proche* on the grammaticalization of [BE *going to* INF] in English.

2.4 The contact hypothesis and contact-induced language change

In this section, I will provide a theoretical framework of the contact hypothesis (Danchev and Kytö 1994; 2002) and of contact-induced language change. The potential influence of contact through translation from French into English on the *go*-future in English has not been studied extensively, yet it is important to take it into account when studying the conventionalization of the [BE *going to* INF] construction in English

(Danchev and Kytö 1994; 2002). First, I will discuss the contact hypothesis and the main points of view treated in the literature as well as some of the arguments proposed by Danchev and Kytö (1994; 2002) supporting this hypothesis. This topic needs to be discussed alongside the theory behind contact-induced language change, which I will discuss thereafter. Lastly, I will summarize the main points adduced in this section.

Given that movement verbs are one of the most prominent sources for futures (Bybee, Perkins, and Pagliuca 1994), and that there was an important translation practice of French texts into English (Danchev and Kytö 1994), there might be reason to assume that French might have influenced the rise of the *go*-future in English to some extent. However, it has been argued that the French influence has been more or less neglected in research on the grammaticalization of the English *go*-future (Danchev and Kytö 1994; Pertejo 1999; Danchev and Kytö 2002). Yet, Danchev and Kytö (1994) emphasize the importance of taking this potential influence into account. In fact, they remark that the rise and spread of the *go*-future in English should no longer be studied in isolation, and stress that further investigation into the *go*-futures in English and French during the Later Middle English and Early Modern English periods might produce some relevant results. To that extent, Danchev and Kytö (1994) provide an extensive discussion of similarities and differences between the two constructions in their respective contemporary forms.

In fact, there is no doubt that the development of the French *go*-future preceded the English one chronologically (Danchev and Kytö 2002). Nevertheless, Danchev and Kytö (2002) emphasize that it is difficult to draw any solid conclusions as to whether the construction would have emerged without the outside (French) influence. The development of [BE *going to* INF] could equally have been the result of independent developments. In fact, as mentioned above, other scholars point out that verbs denoting movement are actually one of the most common sources for future markers (Bybee and Pagliuca 1987). This is confirmed by Danchev and Kytö (1994), who nuance their contact hypothesis by stating that there is nothing remarkable about this grammaticalization process, as a number of languages around the world demonstrate such a gradual transition of motion verbs, *go* in particular, towards auxiliaries of future time reference, and should therefore be considered as a fairly general or near-universal development.

These claims seem to be endorsed in the literature regarding contact-induced language change. Thomason (2003) defines contact-induced language change as the linguistic results of contact between two or more languages.

Heine and Kuteva (2003) provide a theoretical framework for why processes of constructionalization (grammaticalization) and contact-induced language change concurrently contribute to grammatical change. They argue that contact-induced language change follows and is constrained by the universal principles of grammaticalization. This is in line with Thomason (2003), who states that because the actual processes of contact-induced change parallel processes of internally motivated change, the main difference between internally motivated linguistic change and contact-induced language change lies in the sources of the change, not in the change processes themselves. Heine and Kuteva thus define contact-induced grammaticalization as a grammaticalization process that is due to the influence of one language on another, and of which this influence manifests itself in the transfer of linguistic material from one language (the model language), to another (the replica language).

Heine and Kuteva (2005) argue that this transfer across languages is regular, and can often extend over centuries or even millennia. In addition, they note that this type of language change is a very complex process to investigate, as not all stages are necessarily a direct product of language contact. Therefore, it is difficult to determine

whether language contact played a role in the first place, and if so, if it might have just formed a trigger for other changes to occur. Additionally, it could equally be possible that the linguistic change originally did not involve language contact, but is affected by language contact at some later stage.

In addition, Thomason (2003) mentions an important notion in the study of contact-induced language change. They explain that one can only assume that contact between languages has been a source of linguistic change, when that change would have been unlikely or less likely to occur if that specific contact situation had not been present. On that account, they also note that establishing contact as a cause of language change is possible under favorable circumstances, yet impossible under less favorable circumstances. Such favorable or 'easy' cases are those in which both form and function have been adopted from another language, being expressed by transferred morphemes. However, less favorable cases are those in which the change consists of structure alone, expressed by native morphemes. In the latter case, all evidence for the possible interference of one language in another is circumstantial, and it is thus difficult or even impossible to construct a linguistic time line of which change occurred at which time and why.

In conclusion, it is clear that there is substantial reason to assume that French might have had an influence on the conventionalization of the compositional string 'BE *going to* INF' into the construction [BE *going to* INF] as a future marker in English, possibly through translations of French texts into English. However, given that motion verbs are one of the most prominent sources for futures than other verbs or lexical material of any kind (Bybee, Perkins, and Pagliuca 1994), it might be difficult or even impossible to make solid conclusions in this respect according to Thomason (2003).

2.5 Hypotheses

I will study the differences between English translations and non-translated English texts in frequencies of occurrence (measured in overall normalized frequencies per million words, and in normalized frequencies per million words calculated per text) of two *go*-constructions in English – progressive [BE *going to* INF] and non-progressive [GO to V] – as well as of the contexts in which these constructions occur in two distinct time frames in Early Modern English, while also comparing these observations to the French source texts of these translations. Even though the role of frequency has been said to be ill-operationalizable in the study of conventionalization (cf. 2.2.2), Petré (2019) has demonstrated otherwise, and argues for a quantitative analysis of frequency shifts in the contexts in which assemblies occur. In this research, these will be measured by context parameters capturing the main changes that have been discussed in the literature in both of the *go*-futures, on which the collected corpus was manually annotated: *motion, animacy, intentionality, predictiveness*, and *voice*.

Based on the theoretical framework presented above, I formulate the following hypotheses in this research:

1. The following differences will be observed in the *go*-constructions in general:
   (a) In both studied time frames (EARLY and LATER) combined, both *go*-constructions (progressive [BE *going to* INF] and non-progressive [GO to V]) combined will display a higher normalized frequency in translations than in reference texts, indicating that the constructions are generally used more in translations than in reference texts,

which would point toward the possibility of their use being influenced by the presence of *futur proche* in the source texts.

(b) In both text types (translations and reference texts) combined, both *go*-constructions together (progressive [BE *going to* INF] and non-progressive [GO to V]) will display a higher normalized frequency in LATER than in EARLY, indicating that these constructions have become more conventional over time, regardless of the contact with French source texts.

(c) The French source text corpus will display a higher normalized frequency of the *go*-construction than the English corpus (translation and reference corpus combined) in EARLY, but this difference will have narrowed in LATER. This would confirm that the French *go*-construction was already more conventional at that time, and that the English *go*-construction increased in frequency accordingly.

(d) Both *go*-constructions (progressive and non-progressive) combined will display a higher normalized frequency in reference texts in LATER than in EARLY, indicating that both constructions have become more conventional in non-translated language over time. However, the difference in normalized frequency of the *go*-construction in translations between EARLY and LATER will be smaller, indicating that the construction had conventionalized earlier in translations than in reference texts.

(e) The progressive [BE *going to* INF] and non-progressive [GO to V] constructions individually will display a higher normalized frequency in reference texts in LATER as compared to EARLY, indicating that each of the constructions individually have become more conventional in non-translated language over time. However, the difference between EARLY and LATER will be narrower in translations, indicating that the conventionalization of both constructions individually had already set in earlier in translations than in reference texts.

2. Regarding progressive [BE *going to* INF] and non-progressive [GO to V] separately, the EARLY translations will display the following differences from the EARLY reference texts, measured in proportions as well as in normalized frequencies calculated per text:
   (a) Translations will use the *go*-construction more often in motionless contexts than reference texts.
   (b) Translations will use the *go*-construction more often in contexts with non-agentive or inanimate subjects than reference texts.
   (c) Translations will use the *go*-construction more often in contexts where there is no intentionality expressed than reference texts.
   (d) Translations will use the *go*-construction more often in contexts about pure predictions than reference texts.
   (e) Translations will use the construction more frequently in a passive voice than reference texts.

3. The same trends as in (2) will be observable in LATER, but the discrepancy in normalized frequencies between the translations and reference texts will be smaller than in EARLY, indicating that the use of these constructions in

these contexts has increased in LATER compared to EARLY, and therefore, that the construction has become more conventional in non-translated language over time.

4. The same trends as in (2) will be observable in the EARLY source text corpus as compared to the LATER source text corpus.

5. Using binary logistic mixed effects regression, the above discussed parameters (*motion, animacy, intentionality, predictiveness,* and *voice*) as well as the use of progressive or non-progressive *go*-constructions will be significant predictors for whether or not a text is a translation or a reference text in EARLY, but this will no longer be the case in LATER, indicating that the construction has become more conventional in non-translated language over time. In addition, these context parameters will be significant predictors for the time frame (EARLY or LATER) of French source texts.

3. Methodology

3.1 The corpus

In this section, I will first discuss how I collected the corpus (3.1.1), and subsequently how I engineered it by preprocessing, querying, annotating, and manipulating the data (3.1.2). Lastly, the experimental set-up will be discussed in 3.2. The corpus engineering was executed in Python (version 3.9.7), and all analyses were carried out in R (version 4.0.3.) (R Core Team 2020).

3.1.1 Corpus collection. First, two distinct time frames were delineated. It was important to capture a time frame where the development of [BE *going to* INF] as a future marker was still in its infancy, while *futur proche* was arguably already further ahead (cf. 2.3.2), as well as a time frame where the conventionalization of the future semantics in [BE *going to* INF] had become more pronounced. Petré and Van de Velde (2018, 877) argue that the conventionalization of the future meaning of [BE *going to* INF] most likely occurred between 1620 and 1640 (cf. 2.3.1). By way of providing some margin, two time frames of thirty years were selected: one starting at 1580 and extending to 1609, hereafter referred to as EARLY, and one starting at 1680 and extending to 1709, hereafter referred to as LATER. In this way, their respective start and ending dates are exactly one hundred years apart. In addition to that, the first selection starts forty years prior to the lower limit of the assumed conventionalization range of 1620–1640 (Petré and Van de Velde 2018), while the second selection starts forty years after the upper limit of that range. However, due to the rigorous selection criteria which translations, reference texts, and source texts needed to meet, the LATER corpus only extends to 1699. Resultingly, there are two time frames in this corpus: EARLY (1580-1609) and LATER (1680-1699).

For both time frames, twenty translations were selected from the Early English Books Online (EEBO) corpus[1] and Early Modern Multiloquent Authors (EMMA) corpus[2]. Please note that in LATER, two separate translations were traced to one and the same source text. In addition, two different French source texts in LATER were traced to one translation, for which one reference text was selected. In this selection process, the primary concerns were 1) whether the French source text had been traced, and 2) whether the translator had been traced. Only those translations of which the French source text had been traced were eligible for selection. For this information, I fully relied on the metadata provided by EEBO. Subsequently, the corresponding French source text was collected from Frantext[3] corpus (ATILF) or from the Google Books corpus (Google Books). Both corpora provide their texts in XML format. This will be further discussed in 3.1.2. Given the limited availability of translations (and source texts) meeting these requirements, no restrictions were made regarding the genre of the texts at this stage. Consequently, at this point, two subcorpora had been collected, each divided in two time frames (EARLY and LATER): a translation corpus, and a source text corpus.

1 EEBO provides digitized versions of work printed in the British Isles and North America as well as works in English printed elsewhere from 1470-1700. The Text Creation Partnership (TCP) provides these in SGML/XML encoding.

2 EMMA is a large-scaled specialized corpus of 90 million words, that allows for the in-depth analysis of individuals' language use while also taking into account the community-level usage. For more information and details on the compilation and selection criteria of EMMA, please refer to Petré et al. (2019).

3 Frantext is an online corpus of 260 million words, containing a wide variety of French and francophone texts from the 10th until the 21st centuries.

Subsequently, a reference corpus consisting of non-translated English texts was collected. The objective was to find texts of which the authors were the translators of the corresponding translations. However, this was not always possible, for instance when the translator had not been traced, or when the translator in question did not have any non-translated texts. When the translator had been effectively traced by EEBO, the reference text had to be from the same decade as the translation, and it should preferably be of the same genre. When the translator had not been traced or when they did not author a non-translated English text, the reference text needed to be from the exact same publication year as the translation, and of the same genre (although there is one exception included because of availability reasons, where the translation genre is *poetry*, while the reference genre is *religious*).

This selection process resulted in three subcorpora (translation corpus, source text corpus, reference corpus) each consisting of forty texts, divided in two time frames (EARLY of 1580–1609, and LATER of 1680–1699). The entire corpus collected for this research thus consists of 120 unique texts, totalling 10,065,429 words (excluding punctuation). An overview of the number of words per subcorpus as well as their row and column totals can be found in Table 1. In the EARLY translation and reference corpus, all texts were collected from EEBO. In the LATER translation and reference corpus, nine texts were collected from EMMA, and the remaining 31 from EEBO. In the EARLY source text corpus, six texts were collected from Frantext, and the remaining fourteen from Google Books corpus. In the LATER source text corpus, seventeen texts were collected from Frantext, and the remaining three from Google Books corpus. An overview of this distribution can be consulted in Table 2. The distribution of genres across the subcorpora is relatively even. An overview can be found in Table 3 and Figure 1. In addition, an overview of all texts included can be consulted in Appendix 1.

	translation (EN)	reference (EN)	source text (FR)	total
EARLY	2,122,271	1,015,978	2,602,687	5,740,936
LATER	1,744,753	581,559	1,998,181	4,324,493
total	3,867,024	1,597,537	4,600,868	**10,065,429**

Table 1
Overview of number of words (excluding punctuation) per subcorpus, and their respective row and column totals.

	translation (EN)	reference (EN)	source text (FR)
EARLY	EEBO: 20	EEBO: 20	Frantext: 6 Google Books: 14
LATER	EEBO: 15 EMMA: 5	EEBO: 16 EMMA: 4	Frantext: 17 Google Books: 3

Table 2
Overview of the distribution of sources of the subcorpora.

3.1.2 Corpus engineering. In this section, I will discuss how the corpora were preprocessed, queried, and annotated. First, I will discuss how the data collected from EEBO were preprocessed and queried. Subsequently, I will do the same for the texts from EMMA. Lastly, I will explain how I handled the French texts coming from Frantext

	translation (EN)		reference (EN)		source text (FR)	
genre	EARLY	LATER	EARLY	LATER	EARLY	LATER
drama	1	1	1	1	1	1
history	5	3	5	3	5	3
literature	5	5	5	6	5	6
medical	2	0	2	0	2	0
philosophy/morality	2	5	2	5	2	5
poetry	1	1	0	1	1	1
religious	3	5	4	4	3	4
travel	1	0	1	0	1	0
total	20	20	20	20	20	20

Table 3
Overview of the genre distribution across the subcorpora.

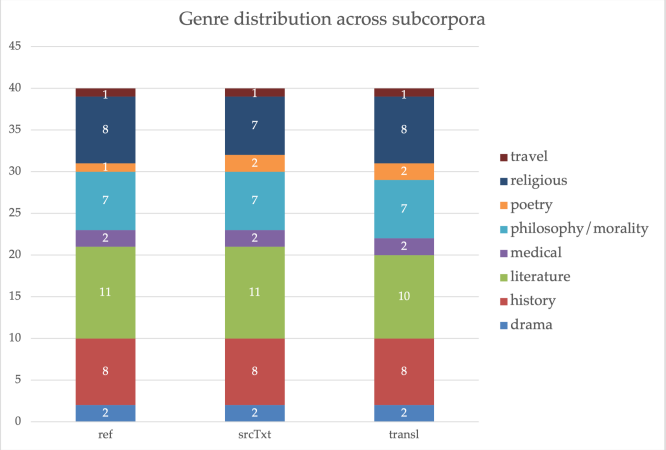

Figure 1
Visualization of the genre distribution across the subcorpora.

and the Google Books corpus. All regular expressions used were adopted from existing regular expressions previously stored in Cosycat (Arévalo and Petré 2017) by Prof. Petré in the context of one of his MA courses and will be added as Appendix 2.

Texts collected from EEBO. The translations and reference texts that were collected from EEBO (cf. Table 2) came in XML format, digitized by Text Creation Partnership (TCP). These were transformed to plain text (*txt*) files using the BeautifulSoup module (Richardson 2007) in Python. Subsequently, they were queried with AntConc software (Anthony 2019) using a regular expression that searches for all progressive instances of the verb *go*, as well as a regular expression that searches for all non-progressive instances of the verb *go* followed by 'to', including historical spellings such as 'vvent' or the use of ſ (the so-called long 's') as an orthographic variant of 's'. These results were exported to a plain text file (*txt*), which was transformed to a data frame in CSV format and manipulated to match the French data frame structure as will be discussed below.

Texts collected from EMMA. The translations and reference texts that were collected from EMMA (cf. Table 2) are available in TEI-XML format. These were queried and annotated using the Collaborative Synchronized Annotation Tool (Cosycat) (Arévalo and Petré 2017). This tool provides a web-based corpus query and annotation interface, and is designed to enable synchronized multi-user annotations of corpora. The selected texts were queried using a regular expression and searches for all instances of [BE *going to* INF] meeting the following conditions:

1. any form of *to be*, including contracted forms and historical spellings (*am, 'm, are, 're, is, 's, art, wert, ...*);
2. typical forms and spellings of *going* (*going, a going, agoing, a-going*);
3. allowing for maximally two words between the copula and the form of *going*;
4. allowing for maximally five words between the form of *going* and *to*.

In addition, the texts were also queried using a regular expression from the same source that searches for all non-progressive instances of the verb *go* as described above. Cosycat stores the annotations in the form of JavaScript Object Notation (JSON) dictionaries. These annotations were extracted into an Excel data frame, which was subsequently manipulated to match the data frame structure of the texts that had been queried using AntConc.

Texts collected from Frantext. The source texts collected from Frantext (cf. Table 2) were provided in XML format. While these transcriptions were generally of excellent quality, the used XML syntax is simply impractical. Instead of storing the information in (child) elements, it stores all of the information in attributes, such as in the following example:

Example sentence: *"Heureusement vous estes philosophe."* (*Entretiens sur la pluralité des mondes*, 1686, De Fontenelle, B.)

```
<line word="Heureusement" pos="ADV" lemma="heureusement"/>
<line word="vous" pos="CLS" lemma="vous"/>
<line word="estes" pos="V" lemma="être"/>
<line word="philosophe" pos="ADJ" lemma="philosophe"/>
<line word="." pos="PONCT" lemma="."/>
```

These attributes were all automatically extracted to a data frame consisting of three columns: *word, POS* (part-of-speech), and *lemma*. The concordancer written for these texts was programmed to build a data frame with the preceding and following context of fifty words for each instance where the lemma is ALLER and the following part-of-speech tag is INF (infinitive).

Texts collected from Google Books corpus. The Google Books corpus provides its automatic transcriptions in EPUB format. This format consists of many files, among which multiple 'content' files containing the textual data stored in XML format. These were subsequently transformed into plain text (*txt*) using the BeautifulSoup module (Richardson 2007) in Python. After tokenization, lemmatization, and parts-of-speech tagging using the spaCy module, a similar data frame of three columns was built: *word, POS* (part-of-speech), and *lemma*. After that, the same concordancer was used as for the Frantext

data, resulting in a data frame with the preceding and following context of fifty words for each instance where the lemma is ALLER and the following part-of-speech tag is VERB.

It should be noted that Google Books uses Optical Character Recognition for the automatic transcription of digitized texts. Therefore, the quality of the transcriptions, and thus the XML data used, is substantially inferior to that of the Frantext data and it could be possible that not everything was transcribed. In addition, the texts were tokenized, lemmatized, and POS-tagged using spaCy's `fr_core_news_sm` module. This is a module trained on high-quality, contemporary French textual data. Evidently, I expect that there will be instances that were incorrectly tagged, and therefore were not extracted by my concordance program. In addition, this module tags verbs in general, and not infinitives in specific. Therefore, the executed query incorporated more noise than when it was able to query only on instances of ALLER followed by an infinitive. These are important limitations to this research, yet inevitable, given the large amount of automatically digitized historical texts and the limited availability of French pretrained tagsets, which are mostly trained on contemporary French.

These preprocessing and querying steps resulted in 1,494 concordances for all English data included, and in 1,363 concordances for all French data included. All of these concordances were manually annotated in Excel, except those retrieved from EMMA, which were annotated in Cosycat. The English concordances were annotated on a multitude of parameters: the type of clause in which the construction is found, motion, intentionality, predictiveness, voice, the position of the *to*-infinitival clause, animacy, the tense of the copula, the grammatical number and person of the copula, and fronting (topicalization of the object of the *to*-infinitive). This was repeated for the French concordances, with the exception of the clause types, and fronting. Evidently, French has a substantially different syntax and these parameters would not be comparable cross-linguistically. In order to avoid any bias in this process, I hid the columns containing the metadata during the annotation stage, so as not to be aware of the time frame or text type of a given text. The resulting data thus consists of two spreadsheets – one for English, one for French – with one row per instance of the respective construction, the preceding context, the following context, a column for each of the parameters discussed in this paragraph, and a variety of metadata. After annotation and removing noise, there were 458 instances of English *go*-constructions and 1336 of the French *go*-construction left.

In addition, these data were manipulated to obtain a number of useful text statistics. First, I obtained the raw number of *go*-constructions in general, of the progressive and non-progressive *go*-construction in specific, and of each of the values in the context parameters *motion, animacy, intentionality, predictivess,* and *voice* per text. Subsequently, these raw frequencies were divided by the word count of the respective text and multiplied by 1,000,000. This produced a data frame containing the raw and normalized (per million words) frequencies of the studied constructions and parameters per text included in the corpus.

In conclusion, the corpus collected for this paper contains 120 texts, divided evenly across three subcorpora (translation, reference, and source text) and two time frames (EARLY and LATER). The English texts were preprocessed and subsequently queried for progressive and non-progressive *go*-constructions, and the resulting concordances were cleaned and annotated on a number of grammatical and semantic parameters. The French texts were preprocessed and then queried on instances of ALLER followed by an infinitive, which were cleaned and annotated on the same grammatical and semantic

parameters, except on clause type and fronting. This resulted in 458 instances in English and 1336 in French, which were used to calculate the raw and normalized frequencies of each of the *go*-constructions per text.

3.2 Experimental set-up

In order to test the first hypothesis, two approaches were used. First of all, the overall normalized frequency of a specific construction in a specific sample was calculated by dividing the raw number of occurrences of that observation by the number of words in that corpus, and multiplying it by 1,000,000. This provides a useful measure of comparing corpora of different sizes, as the resulting number represents a projection of how many times the construction would occur in a similar corpus of a million words. Secondly, as explained above, the normalized frequencies of both *go*-constructions as well as the progressive and non-progressive *go*-constructions separately were calculated in each text, by dividing the raw number of occurrences of the construction by the word count of the text in which it occurred and multiplying by 1,000,000. This approach allows for a more comprehensive interpretation of the data, because outliers influencing the overall normalized frequency can easily be discovered through box plots.

Some context parameters in hypotheses 2, 3, and 4 contain many different values and are rather sparse. To be able to draw more solid conclusions from these, non-binary variables were grouped into binary variables. For example, the variable *animacy* contains values such as *human*, *animal*, *abstract*, or *non-human*. Grouped into *animate* and *inanimate*, the changes in this parameter become more apparent. The distribution of the values of a parameter was measured in terms of percentages taken in the specific sample. For example, when comparing the parameter *animacy* between EARLY translations and EARLY reference texts, the following fictitious table (4) would indicate that 40% of *go*-constructions in EARLY translations occur with inanimate subjects, while only 20% of *go*-constructions in EARLY reference texts occur with inanimate subjects.

animacy	EARLY translations	EARLY reference texts
animate	60.00	80.00
inanimate	40.00	20.00

Table 4
Fictitious table of the distribution of the parameter *animacy* in percentages (%) in EARLY translations compared to EARLY reference texts.

Additionally, normalized frequencies per million words (pmw) of the values that these parameters take in the specific sample were calculated per text. For instance, for the context parameter *animacy*, the normalized frequencies of animate and inanimate subjects in both *go*-constructions combined, in progressive [BE *going to* INF], in non-progressive [GO to V], and in [ALLER INF] were calculated per text. These will be compared across samples using box plots, which are especially useful to obtain a more comprehensive view of the distribution of the data and their potential outliers.

In order to test the last hypothesis, two binary logistic mixed effects regression analyses were carried out to establish whether the parameters of *aspect* (progressive or non-progressive construction), *motion*, *animacy*, *intentionality*, *predictiveness*, and *voice* are significant predictors for whether or not a text is a translation or a reference text in EARLY compared to LATER. In addition, the same regression analysis was carried out for the French source text corpus, to measure whether these context parameters

are significant predictors for the time frame of a text in this corpus. As discussed in the previous paragraph, non-binary variables were grouped into binary variables. The authors were included as a random effect.

Not all variables on which the corpora were annotated were taken into account. Further quantitative and statistical analyses of the clause type, the tense and the grammatical number and person of the copula, as well as of the position of the *to*-infinitival clause and fronting fall outside of the scope of this research. The focus of the analyses carried out in this research will be on the use of *go*-constructions in general, and in particular on the use of the progressive and the non-progressive *go*-constructions respectively, as well as on the contexts in which these occur in terms of the expression of *motion, animacy, intentionality, predictiveness,* and *voice*, as these have been generally discussed to be of particular interest in the grammaticalization of the future semantics of [BE *going to* INF].

4. Discussion of results

In this section, I will discuss the results of the analyses discussed above. I will do so in a hierarchical manner, following the hypotheses as discussed in 2.5. Hypotheses 2, 3, and 4 will be discussed together, because they are related to each other. All normalized frequencies are expressed per million words (pmw).

4.1 General observations in *go*-constructions (Hypothesis 1)

Both *go*-constructions combined indeed occur more than twice as frequently in translations (101.63 pmw) than in reference texts (40.69 pmw). While, based on this observation alone, one could conclude that the use of *go*-constructions in translations is influenced by the presence of a similar construction in the source text, this would be an oversimplication. Yet, considering Figure 2, it becomes clear that, indeed, translations generally display higher normalized frequencies than reference texts in both time frames combined. However, it is also apparent that the difference between the median normalized frequencies in translations (86.37 pmw) and in reference texts (63.11 pmw) is not as substantial as the difference between their respective overall normalized frequencies, given that these are less influenced by the outliers visible in Figure 2. Still, because the interquartile range of normalized frequencies of *go*-constructions in translations is substantially larger than that of reference texts, it can be assumed that the use of *go*-constructions in general is at least more widely spread in translations than in reference texts. However, this could still simply be a translation effect similar to the ones discussed in 2.3.1.

In addition, without distinguishing translations from reference texts, the normalized frequency of both *go*-constructions combined is higher in LATER (93.71 pmw) than in EARLY (76.48 pmw) in general, even though this difference in overall normalized frequency is not as substantial as that between the text types. However, in Figure 3, displaying the distributions of EARLY compared to LATER in normalized frequencies calculated on text level, a similar trend as in Figure 2 can be observed. While the median normalized frequency of texts in LATER (69.89 pmw) is almost equal to that of EARLY (63.40 pmw), the use in LATER seems to be more widely spread than in EARLY, as the interquartile range is noticeably larger in LATER.

As can be seen from Table 5, the normalized frequency of the *go*-construction in the French source text corpus is substantially higher than in the English corpus, both in EARLY as in LATER. From these normalized frequencies alone, it is clear that the *go*-

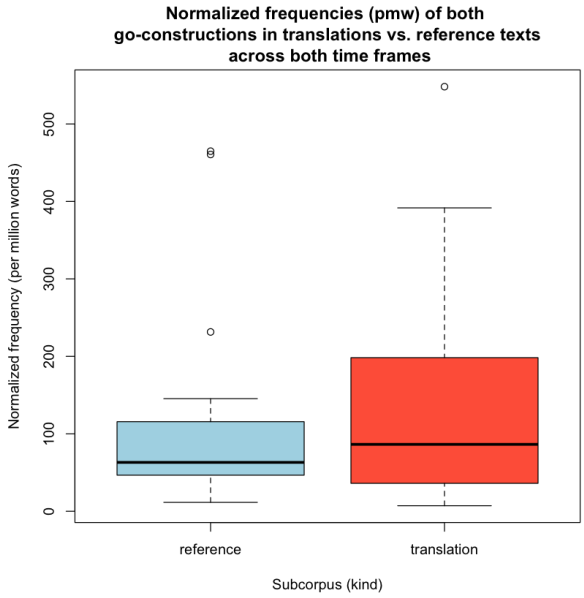

Figure 2
Normalized frequencies (pmw) calculated on text level of both *go*-constructions in translations compared to reference texts across both time frames.

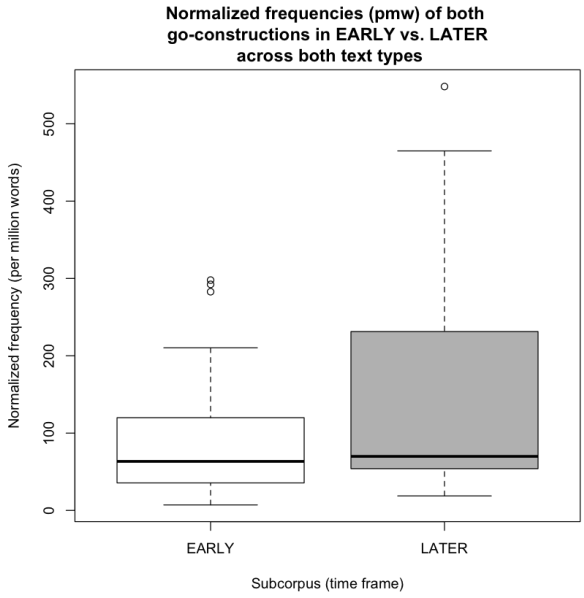

Figure 3
Normalized frequencies (pmw) calculated on text level of both *go*-constructions in EARLY compared to LATER across both text types.

	English corpus	French corpus
EARLY	76.48	334.27
LATER	93.71	233.21

Table 5
Overall normalized frequency (pmw) of *go*-constructions in EARLY and LATER English and French corpora.

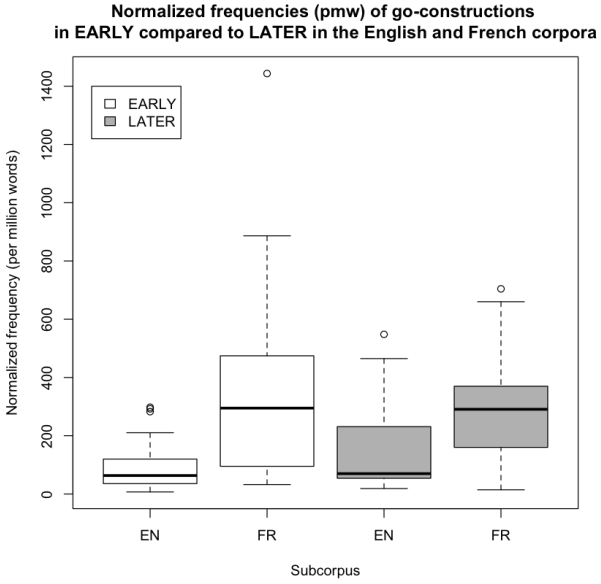

Figure 4
Normalized frequencies (pmw) calculated on text level of the *go*-constructions in EARLY and LATER English compared to EARLY and LATER French

construction was already more conventional in French than in English of the same time frame. Interestingly, the overall normalized frequency is remarkably higher in EARLY source texts than in LATER source texts.

Considering the normalized frequencies calculated per text, the box plot in Figure 4 displays that on average, in EARLY, the use of [ALLER INF] is indeed more frequent and more widely spread than the use of both *go*-constructions in English in general. At the same time, the extreme outlier explains why EARLY French source texts have a higher overall normalized frequency of [ALLER INF] than LATER French source texts. Even though there is a difference of approximately 100 occurrences per million words between EARLY and LATER French source texts in overall normalized frequency, the normalized frequencies calculated on text level effectively paint a more comprehensive picture and are not as influenced by outliers, especially if their medians are considered. More specifically, while the median of the normalized frequencies in EARLY French source texts (294.85 pmw) is still higher than that of LATER French source texts (290.86 pmw), the difference is almost negligible. It appears that the use of [ALLER INF] in LATER French source texts is similar to the use in EARLY, with the exception of the outlier, and the frequency of use actually seems to have come to a standstill over time. This

constitutes yet another indication that, indeed, the use of [ALLER INF] conventionalized earlier in French, and, based on these results alone, its use arguably does no longer stand out. This might result in a new search for unconventional or extravagant forms as language moves along the 'cycle of convention and innovation' (Petré 2016b), but no conclusions can be drawn from these results in that respect.

In LATER, the use of *go*-constructions in English texts has indeed increased to some extent and has become more widely spread, while still remaining substantially less frequently used than the *go*-construction in French source texts of the same time frame.

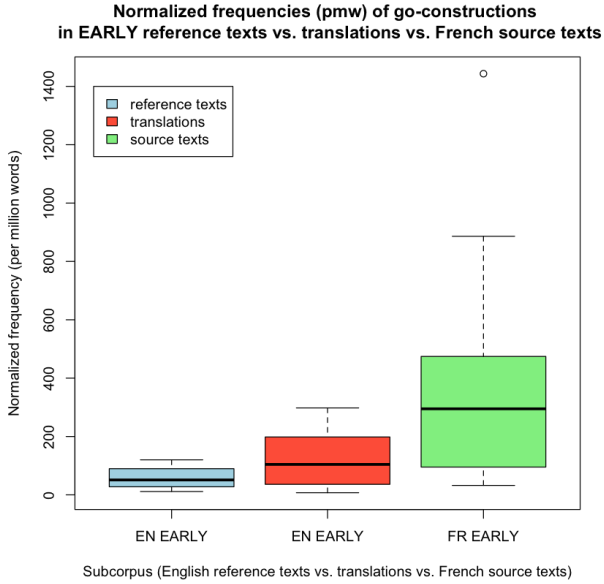

Figure 5
Normalized frequencies (pmw) calculated on text level of the *go*-constructions in EARLY reference texts compared to EARLY translations and EARLY French source texts

At the same time, there is an interesting difference between EARLY and LATER if the text types are considered in this comparison as well. Figure 5 displays that in EARLY, *go*-constructions are indeed most frequently used in French source texts, and are least frequently used in reference texts, while translations seem to be in between the two. These trends suggest an observable influence of French source texts on the use of *go*-constructions in translations in EARLY, as these were not as frequently used in non-translated language of the same time frame. This has changed noticeably in LATER, as can be deduced from Figure 6. This plot shows that the use of *go*-constructions has become more widely spread in LATER reference texts compared to translations. So, while translations contained more *go*-constructions than reference texts in the EARLY time frame, the increase over time in translations is actually less impressive than in reference texts. Since the future semantics of [BE *going to* INF] conventionalized in English somewhere in between these two time frames (Petré and Van de Velde 2018, 877), the increase in frequency in non-translated language in LATER is evident and expected. The higher normalized frequencies in EARLY translations therefore effectively demonstrate that the use of both *go*-constructions was influenced by the presence of French source texts in EARLY.

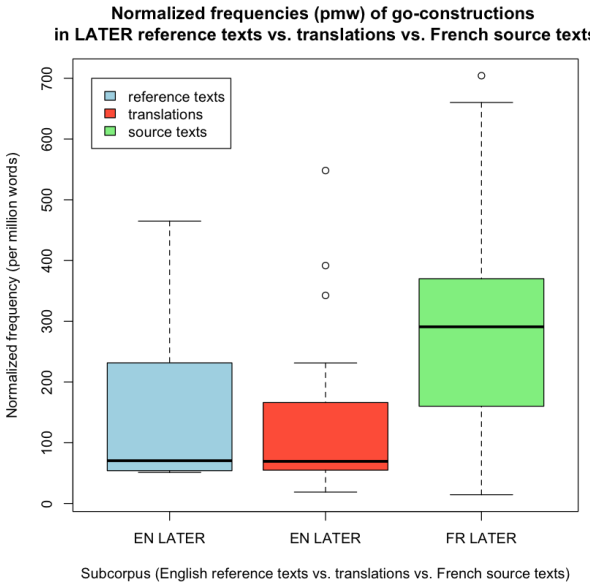

Figure 6
Normalized frequencies (pmw) calculated on text level of the *go*-constructions in LATER reference texts compared to LATER translations and LATER French source texts

	reference texts	translations
EARLY	21.65	102.72
LATER	73.94	100.30

Table 6
Overall normalized frequency (pmw) of both *go*-constructions in reference texts compared to translations in EARLY and LATER.

From the overall normalized frequencies in Table 6, it is clear that there is indeed a substantially steeper increase in frequency over time in reference texts than in translations. In fact, the overall normalized frequency of use of both *go*-constructions in translations actually seems to have stagnated, and has even decreased marginally. This is also reflected in the box plots of the normalized frequencies of both *go*-constructions calculated per text, presented in Figures 7 and 8. Not only are these generally higher in LATER reference texts compared to EARLY reference texts, but also compared to LATER translations. The observation that the frequency of use of *go*-constructions generally has become less widely spread in LATER than in EARLY translations, except for some outliers, is remarkable. Admittedly, it was expected that the difference in normalized frequencies between the two time frames would be smaller in translations than that in reference texts. Yet, if there is an increase in the frequency of use in non-translated language over time, one would logically expect an increase in translations as well. However, these observations could indicate a translation influence, as EARLY and LATER translations seem to follow the tendencies of the *go*-construction in French source texts. In EARLY, the use of *go*-constructions seems to be somewhere in between the frequencies of use of reference texts and source texts (cf. Figure 5). In LATER, the

frequency of use in French source texts has more or less stagnated compared to EARLY, and this is also visible in the LATER translation corpus (cf. Figure 6).

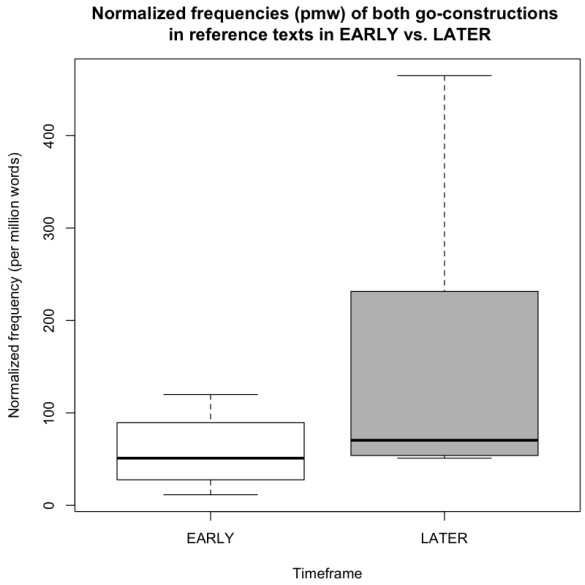

Figure 7
Normalized frequencies (pmw) calculated on text level of the *go*-constructions in reference texts in EARLY compared to LATER.

This stagnation in translations becomes clear when looking at the progressive and non-progressive constructions individually in terms of overall normalized frequencies, provided in Table 7. In reference texts, the progressive construction has become almost seven times more frequent in LATER (44.71 pmw) compared to EARLY (6.89 pmw). In addition, while the non-progressive *go*-construction was already more frequent in EARLY reference texts (14.76 pmw) than the progressive construction, it has not followed the same rapid increase in LATER reference texts (29.23 pmw). In translations, the progressive construction is more frequently used than in reference texts of the same time frame, while the increase in frequency is not as substantial, as was expected. Yet, the stagnation in both *go*-constructions over time in translations observed in Table 6 and in Figure 8 is explained by the decline in the use of non-progressive [GO to V] in translations over time. While non-progressive [GO to V] was used more often in EARLY translations than in reference of the same time, as expected, its frequency of use has declined in LATER translations, while it has increased in reference texts. Still, the use of the non-progressive *go*-construction remains twice as frequently used in LATER translations than reference texts. In one way, the outcome that the difference in normalized frequency between LATER reference texts and translations would become smaller, holds true. However, the decline in translations over time was not expected.

In Figures 9 and 10, it appears that reference texts display a steep increase in the frequency of use of the progressive *go*-construction in LATER, while the use of the non-progressive *go*-construction seems to have stayed more or less the same in LATER. The box plot provides a more nuanced overview of the results than the overall

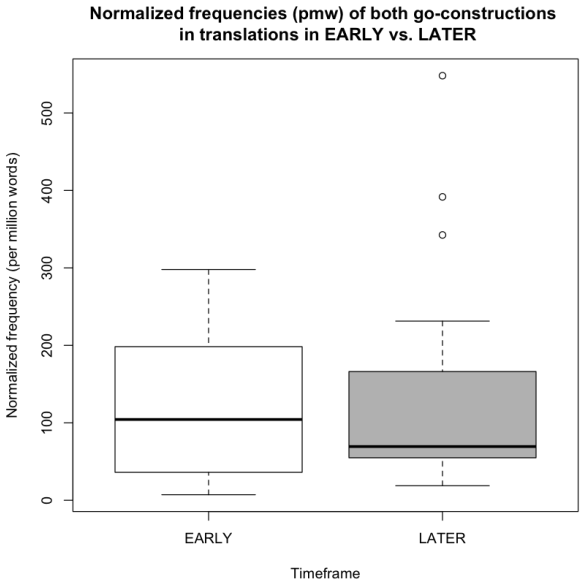

Figure 8
Normalized frequencies (pmw) calculated on text level of the *go*-constructions in translations in EARLY compared to LATER.

	reference texts		translations	
	progressive	non-progressive	progressive	non-progressive
EARLY	6.89	14.76	28.74	73.98
LATER	44.71	29.23	53.30	46.99

Table 7
Overall normalized frequencies (pmw) of progressive and non-progressive *go*-constructions per time frame and per text type.

normalized frequencies as such, since the median normalized frequency of progressive constructions in LATER reference texts remains zero. In translations, the progressive *go*-construction is used substantially more frequently in LATER compared to EARLY. While the interquartile range of normalized frequencies is smaller in LATER translations than in LATER reference texts, the median normalized frequency is substantially higher than that of reference texts, and the use of progressive *go*-constructions in LATER translations therefore seems more widely spread than in reference texts of the same time. The use of non-progressive [GO to V] in translations has generally declined over time. However, from the overall normalized frequency as well as from the normalized frequencies calculated per text, it can be inferred that EARLY translations use the non-progressive *go*-construction more frequently than reference texts of the same time frame, and more frequently than LATER reference texts and translations. Therefore, the presence of [ALLER INF] in French source texts arguably might have given rise to the use of [GO to V] in EARLY translations. This would make sense, given that it is a construction which resembles the French source construction in that it is formed with a motion verb followed by an infinitive, and which was readily available in English, as can be deduced

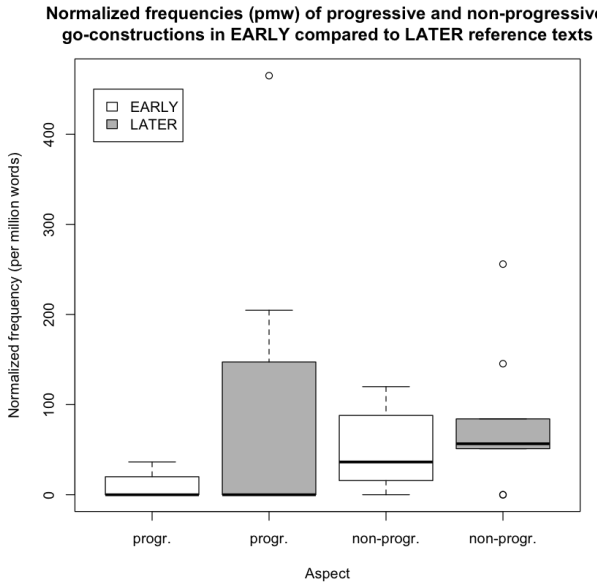

Figure 9
Normalized frequencies (pmw) calculated on text level of the progressive and non-progressive *go*-constructions considered individually in reference texts in EARLY compared to LATER.

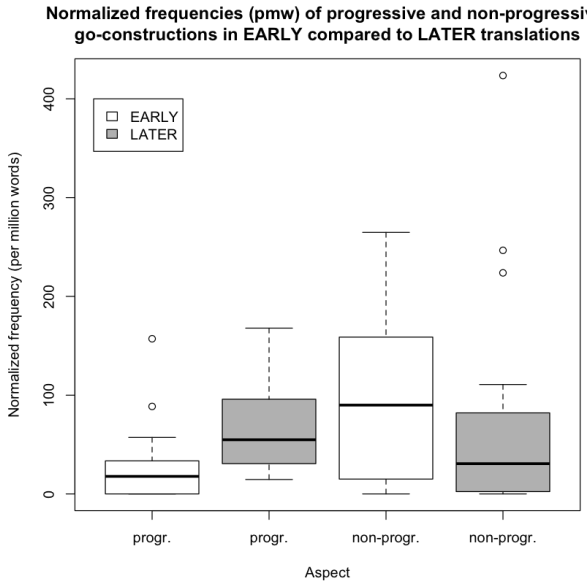

Figure 10
Normalized frequencies (pmw) calculated on text level of the progressive and non-progressive *go*-constructions considered individually in translations in EARLY compared to LATER.

from the frequencies observed in non-translated texts of the same time. The decline in the use of non-progressive [GO to V] in LATER translations is not surprising in that respect, considering that the progressive [BE *going to* INF] construction had become more common in that time, and provided a better alternative for translating [ALLER INF] in the French source texts.

In conclusion, considering that [ALLER INF] is more frequently used than both *go*-constructions combined in English in EARLY, and that translations display a higher frequency of use in EARLY compared to reference texts, it can be concluded that French source texts influenced the use of *go*-constructions in general in EARLY translations. In addition, in LATER, the frequency of use of both *go*-constructions seems to have stagnated or even marginally decreased in translations, which is a trend that is also visible in the French source text corpus. At the same time, however, the frequency of use of both *go*-constructions has increased in LATER reference texts. The influence of French source texts on EARLY translations is especially apparent when it comes to the frequency of use of progressive [BE *going to* INF], as this construction occurs approximately four times more frequently in EARLY translations than in EARLY reference texts. While reference texts display a steeper increase in the use of [BE *going to* INF] than translations in LATER, this increase is arguably due to outliers (cf. Figure 9), as the median normalized frequency of LATER reference texts remains zero. LATER translations show a more gradual increase, but the overall normalized frequency as well as the median normalized frequency of [BE *going to* INF] are higher than in reference texts. This indicates that the use of [BE *going to* INF] is more widely spread in LATER translations compared to reference texts. Yet, the use of [ALLER INF] in French source texts seemingly influenced the use of non-progressive [GO to V] in EARLY translations as well. [GO to V] arguably provided a readily available, resembling option for translating [ALLER INF], and declined in frequency over time due to the rise of progressive [BE *going to* INF].

While these results provide insights into the frequency shifts of *go*-constructions in translations, reference texts, and source texts, they do not yet allow for any conclusions in terms of the conventionalization of the future semantics in the studied *go*-constructions. This will be discussed in the following section.

4.2 Observations in context parameters (Hypotheses 2, 3, 4)

The second, third, and fourth hypotheses investigate context parameters of *motion, animacy, intentionality, predictiveness,* and *voice* in progressive and non-progressive *go*-constructions in EARLY translations and reference texts, LATER translations and reference texts, and EARLY and LATER source texts respectively, in order to estimate the conventionalization of the future semantics in the *go*-constructions. I expect that translations (EARLY and LATER) will display more 'characteristics' of the conventionalization of future semantics in comparison to reference texts of the same time, in that these will use the construction in question more in motionless contexts, with inanimate (or non-agentive) subjects, where no intentionality is expressed, in contexts that are about pure predictions, and in a passive voice. In addition, I expect that the same will hold true for LATER source texts in comparison to EARLY source texts. These assumptions will be measured by investigating the proportions that the values of the parameters take in each respective corpus, as well as the normalized frequencies of these calculated per text, as explained in 3.2.

4.2.1 (Non-)motion. In the process of the conventionalization of futures that originate from a lexical motion verb, an extension to motionless contexts occurs. Considering that

time	*motion*	progressive *go* ref.	progressive *go* tr.	non-progressive *go* ref.	non-progressive *go* tr.	[ALLER INF] source text
EARLY	indeterminate	14.29	4.92	6.67	3.82	5.06
	no	71.43	26.23	26.67	9.55	34.37
	yes	14.29	68.85	66.67	86.62	60.57
LATER	indeterminate	3.85	7.53	5.88	4.88	3.43
	no	50.00	66.67	0.00	4.88	45.71
	yes	46.15	25.81	94.12	90.24	50.86

Table 8
Overview of proportions (%) of the values of the context parameter *motion* in progressive, non-progressive, and French *go*-constructions in translations, reference texts, and French source texts in EARLY compared to LATER.

this occurred in English between the two time frames studied in this corpus (Petré and Van de Velde 2018, 877), and that this was arguably already in place in the French [ALLER INF] construction (De Mulder 2008, 360–361), I hypothesized that translations will use the *go*-construction more often in motionless contexts than reference texts, and that this tendency will be more outspoken in EARLY than in LATER. In addition, it can be expected that EARLY French source texts will use the *go*-construction more often in motionless contexts than EARLY translations and reference texts, indicating an earlier and more advanced stage of the conventionalization of the future semantics. It is also expected that LATER French source texts will display an increase in motionless contexts compared to EARLY French source texts, indicating that this use conventionalized further over time.

From the proportions in Table 8, it appears that the large majority of progressive *go*-constructions in EARLY reference texts expresses non-motion (71.43%) rather than motion (14.29%), while this is not the case in EARLY translations. Interestingly, in reference texts, the proportion of progressive *go*-constructions used in motionless contexts has decreased by 20% over time, while the same proportion increased by 40% in translations over time. This evolution becomes more apparent when considering the normalized frequencies of the progressive *go*-construction in motionless contexts in Figure 11. Please note that one outlier (357.63 pmw)[4] in LATER reference texts was excluded in this box plot, in order to obtain a better scale in between the samples. While it appears from this plot that the progressive *go*-construction is more frequently used in EARLY reference texts than in EARLY translations, it also points out that this use has in fact not decreased in LATER reference texts as compared to EARLY reference texts in terms of frequency per million words, but has rather marginally increased over time. A more impressive increase is visible in LATER translations, which do not only contain substantially more progressive *go*-constructions in motionless contexts than EARLY translations, but also compared to reference texts of the same time.

A similar trend is visible in EARLY in the use of non-progressive [GO to V]. Reference texts (26.67%) in EARLY use the non-progressive construction substantially more often in motionless contexts than translations (9.55%) of the same time frame. However, when considering the normalized frequencies of non-progressive *go*-constructions in Figure

4 Gibbs, Richard. 1687. *The new disorders of love. A gallant novel. Written by Richard Gibbs, of Norwich, philo. medici.*

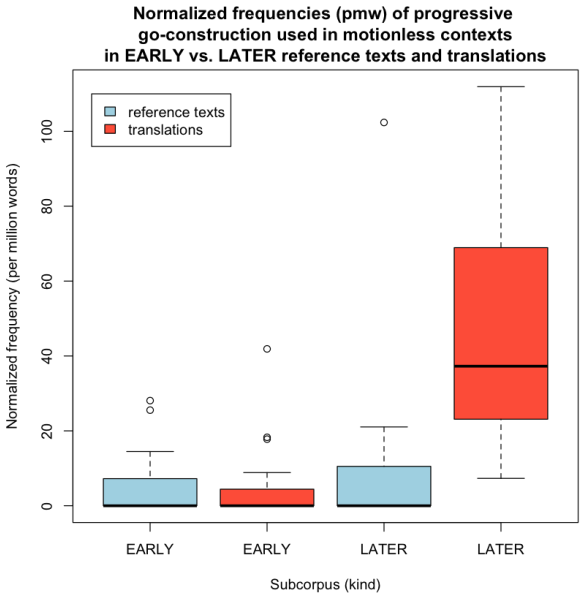

Figure 11
Normalized frequencies (pmw) calculated on text level of the progressive *go*-construction used in motionless contexts in reference texts and translations in EARLY compared to LATER.

12, it becomes clear that this imbalance in proportions shown in Table 8 is due to extreme outliers[5] in EARLY reference texts. This use is actually more widely spread in EARLY translations than in reference texts, and becomes almost obsolete in LATER reference texts and translations.

While, in EARLY as well as in LATER, the majority of [ALLER INF] constructions in French source texts is still used in contexts where motion is expressed, the use of the *go*-construction in motionless contexts has increased by 10% in LATER compared to EARLY. This evolution can also be seen in Figure 13, which confirms that the use of [ALLER INF] in motionless contexts was already more conventional in the EARLY time frame than the use *go*-constructions in motionless contexts in English, and has increased in normalized frequency of use over time, as expected.

In conclusion, in EARLY, reference texts use [BE *going to* INF] slightly more frequently in motionless contexts than translations, while this was not expected. In addition, in LATER, this use has increased slightly in reference texts, but has increased drastically in translations. While this was not what was expected, this observation clearly illustrates that there is in fact an observable difference in the contexts in which [BE *going to* INF] is used in translations, while not yet being suitable in non-translated language of the same time. The use of non-progressive *go*-constructions in motionless contexts is rare, and remains rare throughout the subcorpora. Interestingly, it was used

5 Middleton, Christopher. 1597. *A courtly controuersie, betweene looue and learning Pleasauntlie passed in disputation, betweene a ladie and a gentleman of Scienna. Wherein is no offence offered to the vertuous nor any ill motion to delight the vicious.* (119.92 pmw).
Munday, Anthony. 1581. *The famous historie of Chinon of England with his strange aduentures for the loue of Celestina daughter to Lewis King of Fraunce.* [...] (106.44 pmw).

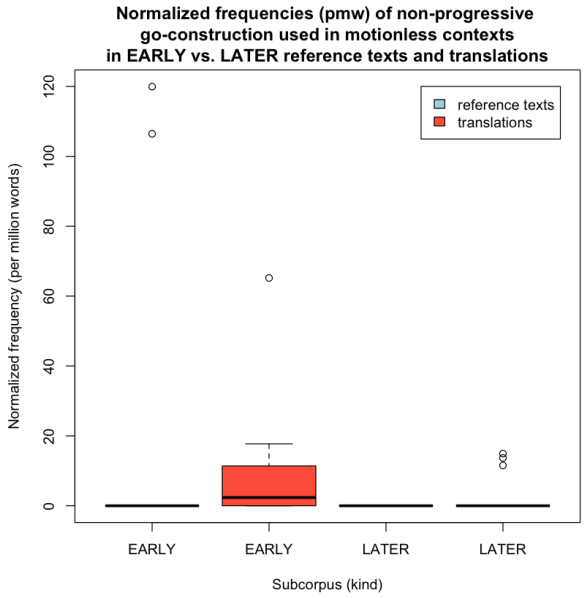

Figure 12
Normalized frequencies (pmw) calculated on text level of the non-progressive *go*-construction used in motionless contexts in reference texts and translations in EARLY compared to LATER.

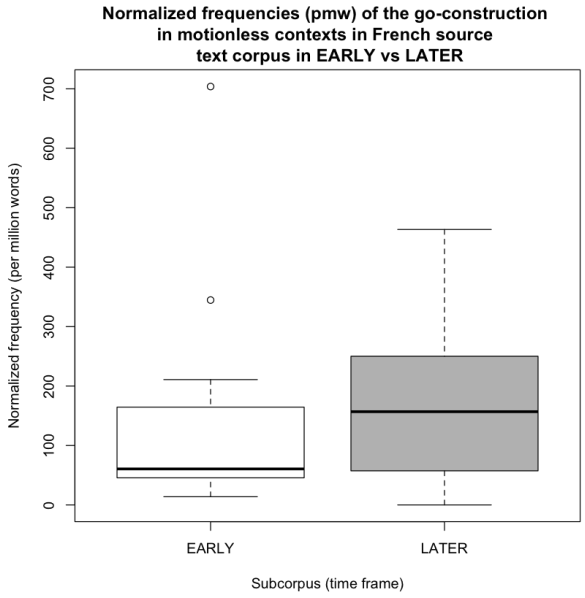

Figure 13
Normalized frequencies (pmw) calculated on text level of the *go*-construction used in motionless contexts in source texts in EARLY compared to LATER.

		progressive *go*		non-progressive *go*		[ALLER INF]
time	*animacy*	ref.	tr.	ref.	tr.	source text
EARLY	indeterminate	14.29	0.00	6.67	5.73	0.92
	animate	85.71	100.00	93.33	94.27	96.32
	inanimate	0.00	0.00	0.00	0.00	1.72
	dummy	NA	NA	NA	NA	1.03
LATER	indeterminate	11.54	3.23	0.00	9.76	0.21
	animate	84.62	90.32	100.00	87.80	89.06
	inanimate	3.85	6.45	0.00	2.44	10.09
	dummy	NA	NA	NA	NA	0.64

Table 9
Overview of proportions (%) of the values of the context parameter *animacy* in progressive, non-progressive, and French *go*-constructions in translations, reference texts, and French source texts in EARLY compared to LATER.

most frequently in EARLY translations in terms of normalized frequencies, but has become almost obsolete in LATER, contrary to what was expected. This might be due to the rise in the use of the progressive *go*-construction in LATER. These analyses also confirm that, as expected and as discussed in 2.3.2, the use of [ALLER INF] in motionless contexts was already more conventional in EARLY compared to both *go*-constructions in English. In addition, its frequency of use has increased over time. While it cannot be concluded that the use of [ALLER INF] in motionless contexts in French source texts has influenced the use of both *go*-constructions in EARLY translations, it is likely that the use of [BE *going to* INF] in motionless contexts in LATER is stimulated by the example set in French source texts.

4.2.2 Animacy. In the gradual process of the constructionalization of the *go*-future in French and in English, the construction is able to be used in contexts with inanimate, non-agentive subjects (De Mulder 2008; Bres and Labeau 2018; Petré 2019). Therefore, reasoning that the conventionalization of the future semantics in [ALLER INF] preceded that of [BE *going to* INF] chronologically, and that the construction thus was able to be used with inanimate subjects earlier, it is expected that translations will use the *go*-construction more often with inanimate subjects than reference texts. In particular, I hypothesized that EARLY translations would use the *go*-construction more often with inanimate subjects than EARLY reference texts, and that this difference would have become smaller in LATER translations compared to LATER reference texts, indicating that the use of inanimate subjects in these constructions initiated in translations, but became more frequently used and thus more conventional in non-translated language over time.

Both in EARLY translations and reference texts, there are no progressive nor non-progressive *go*-constructions used with inanimate subjects. In LATER, however, there is an observable difference. It is clear from the proportions shown in Table 9 that the use of inanimate subjects in progressive *go*-constructions has increased in LATER in general, and in LATER translations in specific, while still remaining rather rare. This is confirmed in Figure 14. LATER translations simply contain more outliers, but the median normalized frequencies of [BE *going to* INF] with an inanimate subject in LATER reference texts and translations are both zero.

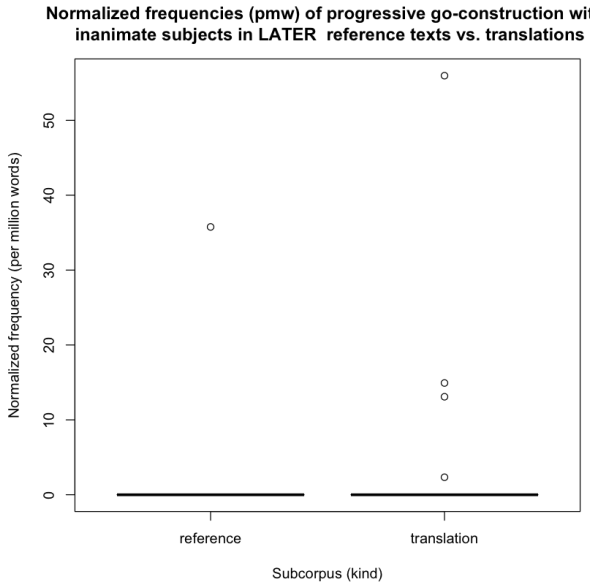

Figure 14
Normalized frequencies (pmw) calculated on text level of the progressive *go*-construction with inanimate subjects in reference texts compared to translations in LATER.

Non-progressive *go*-constructions used with inanimate subjects do not occur in EARLY at all, nor in LATER reference texts. However, a minority (2.44%, n=2) of non-progressive *go*-constructions in LATER translations is used with inanimate subjects. This observation is, however, too small to draw any conclusions.

In the French source text corpus, there is an important increase in the use of inanimate subjects over time. More specifically, 10.09% is used with an inanimate subject in LATER, compared to only 1.72% of EARLY. Figure 15 illustrates that the use of inanimate subjects in [ALLER INF] was already substantially more frequently used in EARLY compared to the use of inanimate subjects in EARLY reference texts and translations. In addition, inanimate subjects in [ALLER INF] have not only become more widely spread over time, but the median normalized frequency of this use has also increased noticeably in LATER compared to EARLY.

In short, while the general hypothesis that progressive *go*-constructions occur more with inanimate subjects in translations than in reference texts holds true to some extent, this is only the case for the LATER subcorpus, as *go*-constructions are not used with inanimate subjects in the EARLY translation nor reference corpus. In addition, these results confirm the hypothesis that [ALLER INF] is used more frequently with inanimate subjects in LATER than in EARLY French source texts. This constitutes supplementary evidence that in terms of animacy of the subject, the future semantics of the *go*-construction had already conventionalized earlier in [ALLER INF] and continued to do so over time. These results point toward a slight influence of French source texts on the use of [BE *going to* INF] in LATER translations, although this conclusion remains tentative in that the outliers might be due to a translation effect.

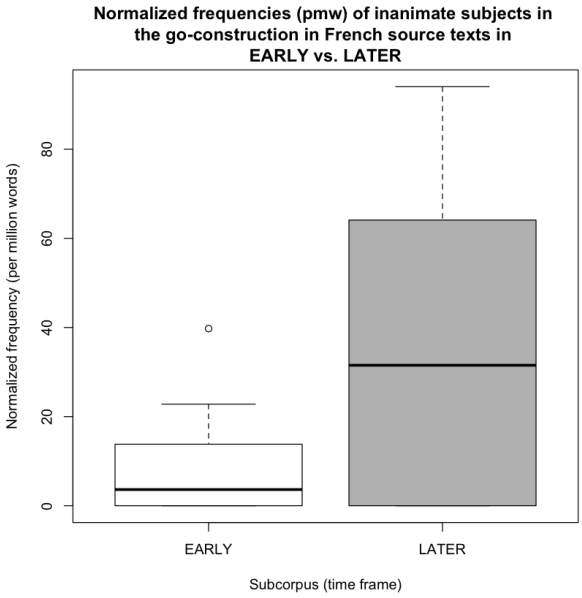

Figure 15
Normalized frequencies (pmw) calculated on text level of the *go*-construction used with inanimate subjects in source texts in EARLY compared to LATER.

4.2.3 Intentionality. In the process of the constructionalization of lexical verbs denoting motion to grammatical future markers, both ALLER and GO lose in spatial meaning, while the expression of intention becomes more prominent to the extent that intention can be expressed without movement. Eventually the subject's intention becomes redundant as well, to the point where the construction merely expresses prediction or future (Bybee, Perkins, and Pagliuca 1994; De Mulder 2008; Disney 2009a). Given that Petré and Van de Velde (2018, 877) argued that the conventionalization of the future semantics of [BE *going to* INF] most likely occurred between 1620 and 1640, which is right in between the two time frames studied in this research, and considering that [ALLER INF] was already combined with infinitives that no longer depended on the subject's intention at least a century earlier (De Mulder 2008, 361–362), I hypothesized that translations use the *go*-construction more often in contexts where no intentionality is expressed than reference texts, both in EARLY and LATER, while the difference between the two text types will be greater in the former than in the latter. In addition, I hypothesized that LATER source texts will use the French *go*-construction more often than EARLY source texts, indicating a further advanced conventionalization of the future semantics in [ALLER INF].

From the proportions in Table 10, it appears that progressive *go*-constructions occur only marginally more frequently in non-intentional contexts in EARLY translations (3.28 %, n=2) than in EARLY reference texts. Even though this difference is negligible, it seems that the non-intentional use of [BE *going to* INF] is influenced by the non-intentional use of [ALLER INF] in the source text. The example given in (13) is an instance of the EARLY translations corpus with a non-intentional use of [BE *going to* INF], which is a word-for-word translation of the source text, given in (14). Both instances were annotated to be cases of non-intentionals.

time	*intentionality*	progressive *go* ref.	progressive *go* tr.	non-progressive *go* ref.	non-progressive *go* tr.	[ALLER INF] source text
EARLY	amb.	0.00	1.64	6.67	1.91	1.95
	int.	100.00	95.08	86.67	97.45	92.07
	non-int.	0.00	3.28	6.67	0.64	5.17
	unclear	0.00	0.00	0.00	0.00	0.80
LATER	amb.	3.85	5.38	5.88	3.66	3.65
	int.	76.92	82.80	94.12	95.12	79.18
	non-int.	19.23	11.83	0.00	0.00	16.95
	unclear	0.00	0.00	0.00	1.22	0.21

Table 10
Overview of proportions (%) of the values of the context parameter *intentionality* in progressive, non-progressive, and French *go*-constructions in translations, reference texts, and French source texts in EARLY compared to LATER.

13. [...] And in very truth it was in him great folly, yea rather madneſſe, to goe about to kill a man that *was going to die*, and he (that went to giue the blow) in the ſame perill.
(Erondelle, P., 1609. *Noua Francia : or The description of that part of Nevv France* [...])

14. [...] Er de verité c'estoit en lui vne grande folie, ou plustot rage, d'aller tuer vn homme qui *s'en va mourir*, & que celui qui veut faire le coup soit en méme peril.
(Lescarbot, M. 1609. *Histoire de la Nouvelle France*)

A larger proportion of progressive *go*-constructions in LATER reference texts (19.23%) is used in contexts where no intentionality is expressed than is the case in LATER translations (11.96%). While there is an increase in this proportion in both text types, it is remarkable that the use of the progressive *go*-construction in non-intentional contexts has increased by almost 20% over a period of one hundred years in non-translated language, while this increase is not as steep in translations. Yet, when looking at the normalized frequencies of progressive constructions used in non-intentional contexts in LATER in Figure 16, it is notable that this use has not increased as drastically in reference texts as would appear from the proportions table, but is rather influenced heavily by outliers. Instead, the use of progressive *go*-constructions in non-intentional contexts seems to be more widely spread in translations rather than in reference texts.

When it comes to non-progressive *go*-constructions, from the proportions presented in Table 10, EARLY reference texts (6.67%, n=1) seem to use these more often in contexts where no intentionality is expressed compared to EARLY translations (0.64%, n=1), although the non-intentional use is generally rare. Yet, this non-intentional use is visibly influenced by the French source text, as can be seen in examples (15) and (16), which are given in 4.2.5 to illustrate the influence of French source texts on the use of passive voice. In non-intentional contexts, EARLY translations and reference texts both display a median normalized frequency of zero. In addition, non-progressive [GO to V] does not occur in non-intentional contexts in LATER.

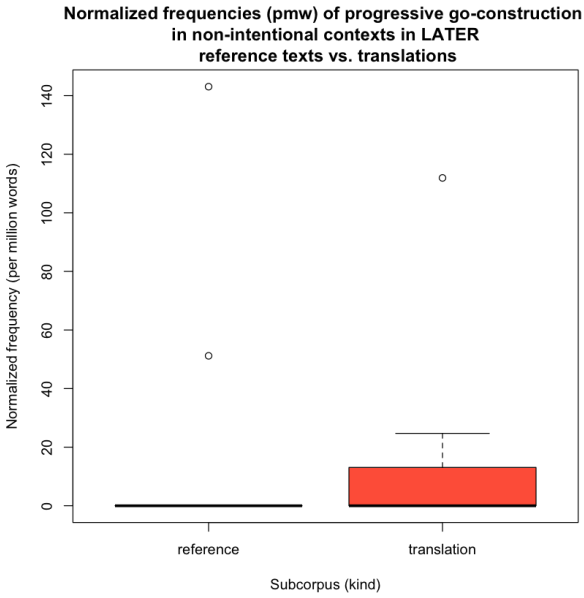

Figure 16
Normalized frequencies (pmw) calculated on text level of the progressive *go*-construction used in non-intentional contexts in reference texts compared to translations in LATER.

In the French source text corpus, the use of [ALLER INF] in non-intentional contexts was already frequent to some extent in EARLY and experienced a steep increase in LATER. This confirms that [ALLER INF] was already at a more advanced stage in the conventionalization of its future semantics in that it could be used in non-intentional contexts, which continued to gain in frequency and conventionalize over time.

In sum, LATER reference texts contain a larger proportion of progressive *go*-constructions used in non-intentional contexts than translations of the same time. In addition, this use has increased in reference texts as well as in translations, although this tendency is stronger in reference texts. Yet, the LATER translation corpus contains texts with consistently higher normalized frequencies of [BE *going to* INF] in non-intentional contexts than the LATER reference corpus, which contains two influential outliers. Furthermore, non-intentional non-progressive *go*-constructions do not occur in LATER, and EARLY reference texts contain a larger proportion than translations of the same time frame, even though this should be nuanced by the fact that the median normalized frequency of use is zero in EARLY reference texts as well as translations. As expected, the future semantics of [ALLER INF] were already common in the EARLY time frame and conventionalized to a higher extent over time. This arguably influenced the use of non-intentionals in EARLY translations, considering that the instances given in (13) and (15, in 4.2.5) consist of a word-for-word translation of a non-intentional use in the French source text.

4.2.4 Predictiveness. During the process of the conventionalization of the *go*-futures, the notion of physical motion through space is weakened and eventually lost, while the new meaning of future or prediction is added (Bybee and Pagliuca 1987; Sweetser 1988; De Mulder 2008; Pérez 1990; Traugott and Trousdale 2013). Therefore, I hypothesized

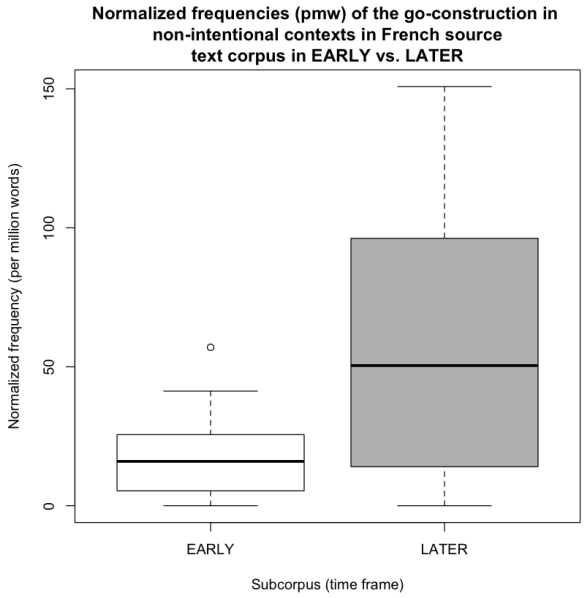

Figure 17
Normalized frequencies (pmw) calculated on text level of the *go*-construction used in non-intentional contexts in source texts in EARLY compared to LATER.

		progressive *go*		non-progressive *go*		[ALLER INF]
time	*predictiveness*	ref.	tr.	ref.	tr.	source text
EARLY	no pred.	100.00	100.00	100.00	99.36	95.98
	pred.	0.00	0.00	0.00	0.64	4.02
LATER	no pred.	96.15	88.17	100.00	97.56	81.55
	pred.	3.85	11.83	0.00	2.44	18.45

Table 11
Overview of proportions (%) of the values of the context parameter *predictiveness* in progressive, non-progressive, and French *go*-constructions in translations, reference texts, and French source texts in EARLY compared to LATER.

that translations will contain more *go*-constructions used in predictive contexts than reference texts, both in EARLY as in LATER, while the difference in occurrence between the two text types would be smaller in LATER than in EARLY. Similarly, I expect LATER source texts to use the *go*-construction more often in contexts about predictions than EARLY source texts.

Table 11 illustrates that while [BE *going to* INF] is not used to make predictions in EARLY, there is an important increase of the progressive *go*-construction being used in contexts which are about pure predictions in LATER, especially in translations. This is confirmed in Figure 18, where the predictive use of [BE *going to* INF] appears to be more frequently used in translations than in reference texts in LATER, while still remaining rather rare.

Non-progressive *go*-constructions are marginally present in contexts about predictions in EARLY translations (0.64%, n=1), and this use increases slightly in LATER

translations, but still remains rare (2.44%, n=2). In addition, reference texts do not use non-progressive [GO to V] to make predictions. The proportions in Table 11 as well as the normalized frequencies calculated per text visualized in Figure 19 illustrate clearly that the use of [ALLER INF] in contexts about predictions was already common, and displays a steep increase over time in LATER.

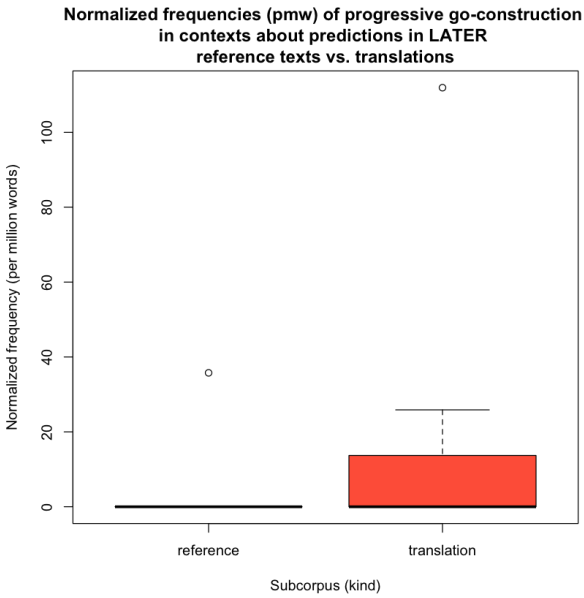

Figure 18
Normalized frequencies (pmw) calculated on text level of the progressive *go*-construction used in contexts about predictions in reference texts compared to translations in LATER.

To conclude, both *go*-constructions increased in their use in contexts about pure predictions over time, although this tendency is stronger for the progressive *go*-construction than for the non-progressive *go*-construction. More specifically, while there is an increase in the use of [BE *going to* INF] in contexts about predictions in LATER reference texts and translations, the increase is more impressive in translations. Given that [ALLER INF] was relatively frequently used to make predictions in EARLY source texts, but *go*-constructions were not in EARLY translations, it must be concluded that French source texts did not influence translations in the EARLY time frame in this respect. Nevertheless, the influence of this use in French source texts becomes clear in the use of [BE *going to* INF] LATER translations, which shows a sharper rise in predictive contexts than reference texts of the same time.

4.2.5 Voice. When *go*-constructions in the process of conventionalization are used with a *to*-infinitive in the passive voice, the notion of agency or control is weakened (Petré and Van de Velde 2018, 875). Considering that the conventionalization of [ALLER INF] chronologically preceded that of [BE *going to* INF], and assuming that there might have been a substantial influence from translations from French into English, I hypothesized that translations will use the constructions more often in a passive voice than reference texts, and that this difference will be greater in EARLY than in LATER. In addition, the French source text corpus is believed to use passive voice more often in EARLY than

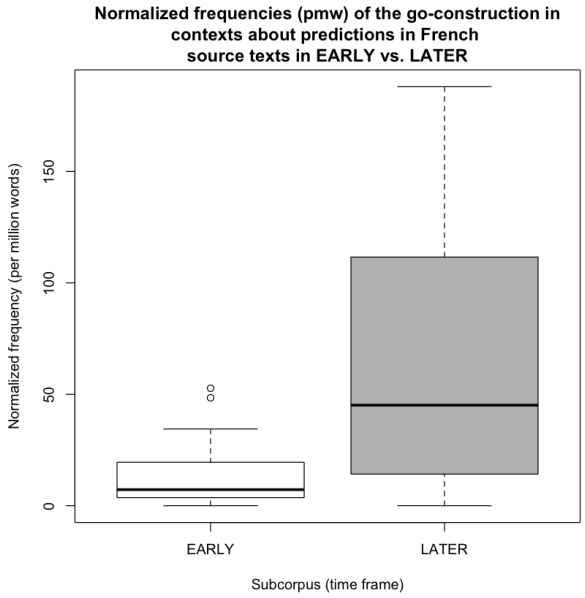

Figure 19
Normalized frequencies (pmw) calculated on text level of the *go*-construction used in contexts about predictions in source texts in EARLY compared to LATER.

		progressive *go*		non-progressive *go*		[ALLER INF]
time	*voice*	ref.	tr.	ref.	tr.	source text
EARLY	active	100.00	96.72	100.00	99.36	100.0
	passive	0.00	3.28	0.00	0.64	0.00
LATER	active	80.77	95.70	100.00	100.00	98.5
	passive	19.23	4.30	0.00	0.00	1.50

Table 12
Overview of proportions (%) of the values of the context parameter *voice* in progressive, non-progressive, and French *go*-constructions in translations, reference texts, and French source texts in EARLY compared to LATER.

translations or reference texts, indicating an earlier more advanced conventionalization. Lastly, a rise in passive voice in LATER French source texts is expected, which would indicate a further conventionalization over time.

In EARLY, passive voice is marginally used in progressive *go*-constructions only in translations (3.28%, n=2) and not in reference texts, while the proportion of [BE *going to* INF] using passive voice in LATER reference texts has increased up to 19.23%, compared to only 4.30% in translations of the same time frame. While the hypothesis that [BE *going to* INF] in EARLY translations will use more passive voice than EARLY reference texts holds true, it is reference texts which display a higher proportion of progressive *go*-constructions with passive voice in LATER (19.23%), rather than translations (4.35%).

This tendency becomes even more apparent when considering the normalized frequencies of [BE *going to* INF] using passive voice calculated per text, as shown in Figure 20. While translations contain two outliers, reference texts have increased drastically in

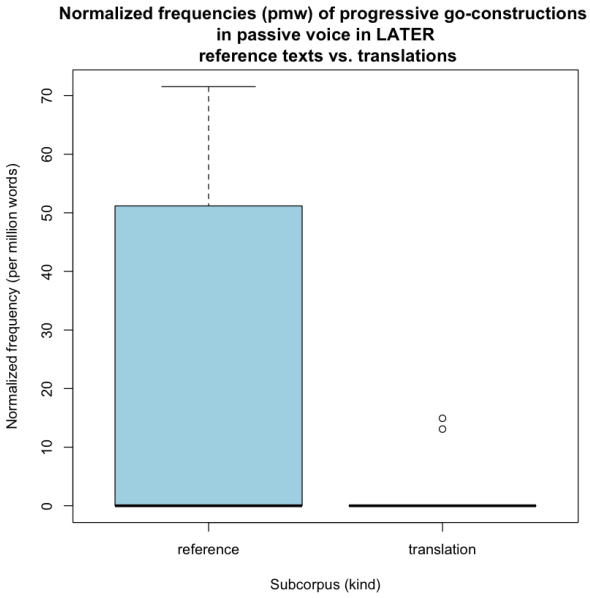

Figure 20
Normalized frequencies (pmw) calculated on text level of the progressive *go*-construction used with passive voice in reference texts compared to translations in LATER.

their use of the progressive *go*-construction in LATER. Admittedly, it cannot be assumed that the use is conventional or widely spread in LATER reference texts, as the median normalized frequency remains zero.

The use of [GO to V] with a *to*-infinitive in passive voice is negligible. While it is remarkable that it only occurs in EARLY translations (0.64%), this should be nuanced by saying that it regards one single instance in the corpus, given in (15).

15. [...] not long since Iohn Baptist said vnto you, behold the lambe of God which taketh away the sinnes of the world: and of whom I my selfe say vnto you at this present: That I *go to be deliuered* vp to death for you; that you shall haue from henceforth my flesh to eat, and my bloud to drinke, for the remission of your sinnes, and for the nou | rishing [...] (Anonymous translator, *Fovvre bookes, of the institution, vse and doctrine of the holy sacrament of the Eucharist in the old Church*, 1600.

It becomes clear that this instance is substantially influenced by the, almost word-for-word, translation of the French source text, which uses the infinitive following the *go*-construction in passive voice as well (16):

16. [...] duquel vous disoit n'agueres leã Baptiste, V viil'Agneau qui oste les Pechez du mõde : duquel ie vous di moi-mêsme maintenāt; Que *ie m'en vai estre liuré à la mort* pour vous; qvous aurez desormais à māger ma chair, & boire mon sang,én renvil fion de vos pechez; en nourriture de vos aines en vie eternelle [...]

(Mornay, Philippe, *De l'institution, usage, et doctrine du sainct sacrement de l'Eucharistie, en l'eglise ancienne*, 1598.)

Unfortunately, while there is a clear case of [ALLER INF] in this example, it has not been extracted by the query. Presumably, this is due to the use of a contemporary tagging module, as mentioned above, and the queries might have missed some concordances because of this. Even though it would be nearly impossible to have a corpus of this size manually checked and queried, this constitutes an important limitation to this research. Still, when looking into the source text corpus in more detail, it becomes clear that the use of passive voice is simply not common in *go*-constructions in general[6], even in texts coming from the Frantext corpus, which is generally of excellent quality and has not been tagged by the contemporary tagger with which the texts from Google Books were tagged. To be more precise, in the French source text corpus as well, in EARLY and in LATER the use of passive voice is rare: it does not occur in EARLY source texts, and takes only 1.5% (n=7) in LATER.

In sum, [BE *going to* INF] is used more often with passive voice in translations than in reference texts in EARLY, while this tendency is inversed in LATER. In non-progressive constructions, the use of passive voice is negligible, but is influenced by translations out of French. The use of passive voice in source texts seems equally rare, even though this might be an artificial effect possibly brought about by the use of a contemporary tagger on historical texts. Therefore, this conclusion remains more tentative, but the results indicate that while the use of passive voice in [ALLER INF] might have influenced word-for-word translations, there does not seem to be a general trend of influence of French source texts in the use of passive voice in either of the *go*-constructions.

In conclusion, the influence of French source texts on translations is most apparent in the use of [BE *going to* INF] in LATER translations. More specifically, LATER translations use [BE *going to* INF] substantially more frequently in motionless contexts compared to reference texts of the same time, and this use might have been stimulated by the example set in French source texts. In addition, French source texts potentially influenced the use of [BE *going to* INF] with inanimate subjects in LATER translations slightly, although this is most likely a translation effect rather than a general trend. Moreover, LATER translations seem to use [BE *going to* INF] more frequently in contexts about predictions than reference texts of the same time. Furthermore, while the use of [ALLER INF] in non-intentional contexts is relatively frequent in EARLY and even more so in LATER, its influence is most apparent in the exceptional non-intentional use of [BE *going to* INF] and [GO to V] in word-for-word translations in EARLY, even though it remains tentative to draw a general trend. Lastly, a similar observation can be made for the use of passive voice in [GO to V] in EARLY, which is noticeably influenced by the translation of the French source text.

4.3 Binary logistic mixed effects regression analyses (Hypothesis 5)

It is clear from the previous results that there are interesting trends to be seen in the use of progressive and non-progressive *go*-constructions in terms of frequency, as well as in terms of the contexts in which these occur. Therefore, this research aims to quantify

6 In raw numbers, only 7 out of 1336 concordances in the entire French source text corpus were used with passive voice.

these tendencies as predictors in a binary logistic mixed effects regression model, in order to verify if and which parameters are significant in predicting whether a given text is a translation or reference text in EARLY, as well as in LATER. Since one can assume that the conventionalization of the future semantics in [BE *going to* INF] has taken place in a period right in between the time frames of the corpora collected for this research, I hypothesized that these parameters would be significant predictors in EARLY, but not in LATER.

To this extent, three binary logistic mixed effects regression analyses were carried out. The first two, one for each of the time frames, have the text type as response variable, the parameters discussed above, and on the use of progressive or non-progressive *go*-constructions (*aspect*) as fixed effect variables, and the author as a random effect. The third one, for the source text corpus, has the time frame as response variable, the context parameters discussed above as fixed effect variables, and the author as a random effect. However, none of these were able to converge.

Motion is the only significant predictor for whether a text is a translation or reference text in the EARLY ($p = 0.030$) English corpus, while the rest of the predictors are not significant ($p > 0.05$). In the LATER English corpus, *voice* is the only context parameter that is a significant predictor for text type ($p = 2.439E-04$). In the French source text corpus, the *intentionality* parameter is the only significant predictor for time frame ($p = 4.798E-03$). Still, it should be noted that these results are not reliable, given that the regression analyses were not able to converge. In addition, these results do not seem in accordance with the trends observed in the context parameters above. This is most likely due to the fact that there are many parameters to take into account, some of which simply do not have enough occurrences in the data for the model to draw solid conclusions. This is an important limitation, and should be accounted for in future research.

5. Limitations

While this research has provided some useful and interesting insights with respect to the conventionalization processes of the *go*-futures in English and in French, it also had some important limitations.

First of all, I relied fully on the metadata provided by EEBO for the corpus collection. This means that if a source text of a translation had been traced by EEBO, it was considered to be the true source text of that translation, and no further restrictions were placed. While, as can be deduced from examples (13) and (15), it is clear that the text in question is the translation of the source text, it was assumed that this is the case for all of the translations and source texts included, but this was not manually checked for each text.

Secondly, some source texts were obtained from the Google Books corpus. The digitizations of these texts were of inferior quality compared to the ones obtained from Frantext. In addition, to be able to query the texts and to use the same concordance program that had been written for the Frantext texts, these were automatically tokenized, lemmatized, and tagged using spaCy's `fr_core_news_sm` module, which is trained on contemporary French data. The inferior quality of the transcriptions combined with the fact that a contemporary French tagger was used on historical texts, could easily have as a result that the concordance program missed cases of [ALLER INF]. This is illustrated by the fact that example (15) was not extracted from the text and therefore is not part of the analyses.

Additionally, while the corpora were annotated on a multitude of parameters containing many different values, not all of these parameters were taken into account in this research, due to its limited nature. Studying the evolution of parameters such as the clause types in which the constructions are used, of the tense and grammatical number and person of the copula, or the evolution of the variety of infinitives that can follow the *go*-future might produce interesting results and I encourage future research to enlarge the collected corpus and analyze these parameters as well. In addition, parameters containing more than two values were made binary, as this allowed to draw more solid conclusions. For example, the variable *animacy* contained values such as *personification, animal, abstract,* or *divine*, but this resulted in rather sparse proportions with only a handful of observations for some values. To be able to draw more solid conclusions as to whether more inanimate subjects were used over time, the values were clustered into *animate* and *inanimate*. While this produced interesting results, it would be useful to enlarge the corpus and analyze such parameters in more detail.

Lastly, the binary logistic mixed effects regression analyses that were carried out in this research were unsuccessful in that they were not able to converge on the data. This is most likely due to the fact that there are many parameters to take into account, for which there are sometimes not enough occurrences in the data to draw solid conclusions from. A straightforward solution for this would be to increase the data substantially, as the simply rendering each parameter binary did not facilitate the convergence of the analyses.

6. Conclusion

This research aimed to assess the relation between the conventionalization of the *go*-futures in English and in French through translations from French into English. While the development of the *go*-future has been studied extensively, its relation to the development of the French *go*-future has not received much attention. However, both developed from a lexical motion verb into a motionless future marker along similar patterns. Given that the development of the French [ALLER INF] preceded the development of [BE *going to* INF] chronologically, and given the translation practice from French into English, it is evident that these developments should no longer be studied in isolation (Danchev and Kytö 1994). The current paper aimed to fill that exact need.

The results point toward an important influence of French source texts on their corresponding English translations, in terms of normalized frequencies of occurrence of the constructions themselves as well as in terms of frequency shifts in the contexts in which these occur. More specifically, French source texts influenced the use of *go*-constructions in EARLY translations in general, and of [BE *going to* INF] in EARLY translations in particular. While reference texts show a sharper rise of [BE *going to* INF] in LATER, the more gradual increase of this construction in LATER translations points toward a more widespread use compared to reference texts. In addition, it seems that the use of [GO to V] was influenced by [ALLER INF] in EARLY translations as well, as it arguably provided a readily available, resembling option for translating [ALLER INF] when [BE *going to* INF] was not yet common, and consequently declined in frequency over time due to the rise of the progressive [BE *going to* INF]. French source texts seem to have influenced the extension to contexts in which *go*-constructions were able to be used in translations as well, when these contexts were not yet as common in non-translated language of the same time. To be more precise, French source texts seem to have stimulated the use of [BE *going to* INF] in motionless contexts substantially in LATER translations, and the use of [BE *going to* INF] with inanimate subjects and to make predictions in LATER

translations to a lesser extent. In addition, there is an observable translation effect in EARLY, where the non-intentional uses of [BE *going to* INF] and [GO to V], as well as the use of [GO to V] with passive voice, are the result of word-for-word translations. Nevertheless, no solid conclusions can be drawn with regard to the influence on non-translated language, as the development of lexical motion verbs into future markers is quite common, and might not be due to the contact with another language.

This research has pointed out that French source texts influenced translations in the use of *go*-constructions in general, [BE *going to* INF] and [GO to V] in particular, as well as in the extension to contexts which enforce the future semantics in [BE *going to* INF]. I conclude that the development of [BE *going to* INF] to a future marker most likely was not initiated by the contact with the French *go*-future, but that the latter probably accelerated its conventionalization process through translations, which in turn might have stimulated the processes of neoanalysis and analogization in English language users. Still, some questions have remained unanswered and some new questions have arisen, so it would be useful to extend the research on this corpus, to collect more data, and to provide more detailed insights into other context parameters that fell beyond the scope of this research. Lastly, as mentioned in 2.2.2, the role of frequency has been argued to be ill-operationalizable. However, the approach proposed by Petré (2019) of analyzing frequency shifts in the contexts in which constructions occur, has proven useful in this research and effectively operationalizes frequency analysis.

7. Future research

As already mentioned, I encourage future research to continue working on the collected corpus and to take into account a number of parameters which fell beyond the scope of this research. In that regard, it will be useful to enlarge this corpus to be able to draw more solid conclusions about these parameters in more detail. In addition, this corpus was specifically designed to compare two distinct time frames, and restrictions were placed on the corpus collection in that way. It would, however, be useful to design the corpus per decade and enlarge the corpus with that objective in mind. This would allow for a more gradual analysis of the evolution over time.

Furthermore, while this research has shed some light on the influence of French source texts on the development and the use of [BE *going to* INF] in translations, it has also raised some new questions on the influence on the use of non-progressive [GO to V]. A more detailed analysis of the development and the use of this construction with regard to translations might prove useful in the future.

In addition, I believe that this research could extend well to translation studies in general, and to research on machine translation in specific. A brief exploration on DeepL Translator (DeepL Translate) pointed out that eight out of ten random French sentences containing *futur proche* were translated into simple future in English rather than a *be going to*-future. In my opinion, it would be interesting to explore this further in contemporary English and French, and to measure quantitatively in which contexts such a machine translation would translate *futur proche* to its English equivalent *go*-future, and in which contexts it resorts to simple future.

8. Acknowledgements

First and foremost, I would like to thank my thesis supervisor, Prof. dr. Peter Petré. This thesis would not have been possible without his accurate feedback and his motivating guidance. Never once has his office door been closed when I visited, and this serves as an excellent metaphor for his supervision during the past year. Thank you.

In addition, I would like to thank Dr. Lisa Hilte for being such a kind and empathetic person in every way, for her motivating words, and for helping me figure out how to manage some aspects of my data. The positive impact you have on students and the efforts you make for them never go unnoticed.

References

Andersen, Henning. 2001. *Actualization and the (uni)directionality of change.* In *Andersen, H. (ed.), Actualization: Linguistic Change in Progress*, pp.225–248. Amsterdam: John Benjamins Publishing Company.

Anthony, Laurence. 2019. Antconc (Version 3.5.8.) [Computer software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software.

Anttila, Raimo. 2003. *Analogy: The Warp and Woof of Cognition.* In *Joseph, B. and Janda, R. (eds.), The Handbook of Historical Linguistics*, pp.425–440. Oxford: Blackwell Publishing.

Arber, Edward. 1869. *The Revelation to the Monk of Evesham: 1196. Carefully edited from the unique copy, now in the British Museum, of the edition printed by William de Machilinia about 1482*. English Reprints 18. London: Edward Arber.

Arévalo, Enrique Manjavacas and Peter Petré. 2017. Enabling annotation of historical corpora in an asynchronous collaborative environment. In *Proceedings of the 2Nd International Conference on Digital Access to Textual Cultural Heritage*, DATeCH2017, pages 9–14, ACM, New York, NY, USA.

ATILF. *Base textuelle Frantext* (Online). ATILF-CNRS & Université de Lorraine. 1998-2021. https://www.frantext.fr/.

Bres, Jacques and Emmanuelle Labeau. 2018. Des constructions de aller et de venir grammaticalisés en auxiliaires. *Syntaxe et sémantique*, (1):49–86.

Budts, Sara. 2014. The grammaticalization of *be going to* in Late Modern English: A corpus-based analysis. Master's thesis, Katholieke Universiteit Leuven.

Budts, Sara and Peter Petré. 2016. Reading the intentions of be going to. On the subjectification of future markers. *Folia Linguistica*, 50(Historica-vol-37).

Bybee, J. 2003. *Mechanisms of Change in Grammaticization: The Role of Frequency.* In *Joseph, B. and Janda, R. (eds.), The Handbook of Historical Linguistics*, pp.602–623. Oxford: Blackwell Publishing.

Bybee, Joan. 2015. *Language change*. Cambridge University Press.

Bybee, Joan L and William Pagliuca. 1987. The evolution of future meaning. In *Papers from the 7th international conference on historical linguistics*, volume 109, pages 109–122, John Benjamins Amsterdam.

Bybee, Joan L., Revere Dale Perkins, and William Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: The University of Chicago Press.

Danchev, Andrei and Kytö. 1994. *The construction be going to + infinitive in Early Modern English.* In *Kastovsky, D. (ed.), Studies in Early Modern English*, pp.59–78. Berlin: Mouton de Gruyter.

Danchev, Andrei and Merja Kytö. 2002. The go-futures in english and french viewed as an areal feature. *NOWELE. North-Western European Language Evolution*, 40(1):29–60.

De Mulder, Walter. 2008. Grammaticalisation, métonymie et pertinence. In *Congrès Mondial de Linguistique Française*, page 028, EDP Sciences.

De Navarre, Marguerite. 1559. *L'heptameron*.

De Smet, Hendrik. 2009. Analysing reanalysis. *Lingua*, 119(11):1728–1755.

De Wit, Astrid, Peter Petré, and Frank Brisard. 2020. Standing out with the progressive. *Journal of linguistics*, 56(3):479–514.

DeepL Translate. Date accessed: 1 January 2022. Available at https://www.deepl.com/en/translator.

Disney, Stephen J. 2009a. The Grammaticalisation of "Be Going To". *Newcastle Working Papers in Linguistics*.

Disney, Steve. 2009b. A domain matrix view of the uses and development of be going to+ infinitive. In *Papers from the Lancaster University Postgraduate Conference in Linguistics & Language Teaching*, page 24.

EEBO. *Early English Books Online*. 1470–1700. Early English Books Online. Digitized by Text Creation Partnership. Online: https://www.proquest.com/eebo and https://textcreationpartnership.org/tcp-texts/eebo-tcp-early-english-books-online.

Elsness, Johan. 1994. On the progression of the progressive in early Modern English. *ICAME journal*, 18(5):5–25.

Fischer, Olga. 2007. *Morphosyntactic change: Functional and formal perspectives*, volume 2. Oxford University Press on Demand.

Google Books. Online: https://books.google.com/.

Gougenheim, Georges. 1929. *Etude sur les périphrases verbales*. Paris: Librairie A.-G. Nizet.

Haiman, J. 1994. *Ritualization and the development of language.* In *Pagliuca, W. (ed.), Perspectives on Grammaticalization*, pp.3–38. Amsterdam: John Benjamins Publishing Company.

Haspelmath, Martin. 1999. Why is grammaticalization irreversible? 37(6):1043–1068.
Heine, B. 2003. *Grammaticalization.* In *Joseph, B. and Janda, R. (eds.), The Handbook of Historical Linguistics*, pp.575–601. Oxford: Blackwell Publishing.
Heine, B., U. Claudi, and F. Hünnemeyer. 1991. *From cognition to grammar: Evidence from African languages.* In *Traugott, E.C. and Heine, B. (eds.), Approaches to grammaticalization*, pp.149–188, volume I. Oxford: Blackwell Publishing.
Heine, Bernd and Tania Kuteva. 2003. On contact-induced grammaticalization. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, 27(3):529–572.
Heine, Bernd and Tania Kuteva. 2005. *Language Contact and Grammatical Change*. New York: Cambridge University Press.
Hoffmann, Sebastian. 2004. *Are low-frequency complex prepositions grammaticalized?* In *Lindquist, H. and Mair, C. (eds.), Corpus approaches to grammaticalization in English*, pp.171–210. Amsterdam, Philadelphia: John Benjamins Publishing Company.
Hopper, Paul J. and Elizabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge University Press.
Hopper, Paul J. and Elizabeth Closs Traugott. 2003. *Grammaticalization*, 2nd edition. Cambridge University Press.
Langacker, R.W. 1977. *Syntactic Reanalysis.* In *Li., C.N. (ed.), Mechanisms of Syntactic Change*, pp.57–139. Texas: University of Texas Press.
Meillet, Antoine. 1912. L'évolution des formes grammaticales.
Mossé, Fernand. 1938. *Histoire de la forme périphrastique être+ participe présent en germanique 1 2*. Paris: C. Klincksieck.
Nesselhauf, Nadja. 2007. The spread of the progressive and its 'future' use. *English Language & Linguistics*, 11(1):191–207.
Pérez, Aveline. 1990. Time in motion: Grammaticalisation of the *be going to* construction in english. *La Trobe University Working Papers in Linguistics*, 3:49–64.
Pertejo, Paloma Nuñez. 1999. Be going to + infinitive: Origin and development. some relevant cases from the helsinki corpus. *Studia Neophilologica*, 71(2):135–142.
Petré, Peter. 2016a. Grammaticalization by changing co-text frequencies, or why [BE V*ing*] became the 'progressive'. *English Language & Linguistics*, 20(1):31–54.
Petré, Peter. 2016b. Unidirectionality as a cycle of convention and innovation: Micro-changes in the grammaticalization of [be going to inf]. *Belgian Journal of Linguistics*, 30(1):115–146.
Petré, Peter. 2017. The extravagant progressive: an experimental corpus study on the history of emphatic [BE V*ing*]. *English Language & Linguistics*, 21(2):227–250.
Petré, Peter. 2019. How constructions are born. the role of patterns in the constructionalization of be going to INF. In *Patterns in language and linguistics*. De Gruyter Mouton, pages 157–192.
Petré, Peter, Lynn Anthonissen, Sara Budts, Enrique Manjavacas, Emma-Louise Silva, William Standing, and Odile AO Strik. 2019. Early modern multiloquent authors (emma): Designing a large-scale corpus of individuals' languages. *ICAME journal*, 43(1):83–122.
Petré, Peter and Freek Van de Velde. 2018. The real-time dynamics of the individual and the community in grammaticalization. *Language*, 94(4):867–901.
Python. Python software foundation. python language reference, version 3.9.7. Available at http://www.python.org.
R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
Richardson, Leonard. 2007. Beautiful soup documentation. *April*.
spaCy. spacy usage documentation. Accessed: 2021-04.
Sweetser, Eve E. 1988. Grammaticalization and semantic bleaching. In *Annual Meeting of the Berkeley Linguistics Society*, volume 14, pages 389–405.
TCP. *Text Creation Partnership*. Online: https://textcreationpartnership.org/.
Thomason, Sarah Grey. 2003. *Contact as a Source of Language Change.* In *Joseph, B. and Janda, R. (eds.), The Handbook of Historical Linguistics*, pp.687–712. Oxford: Blackwell Publishing.
Traugott, E.C. and B. Heine. 1991. *Introduction.* In *Traugott, E.C. and Heine, B. (eds.), Approaches to grammaticalization*, pp.1–14, volume I. Oxford: Blackwell Publishing.
Traugott, E.C. and E. König. 1991. *The semantics-pragmatics of grammaticalization revisited.* In *Traugott, E.C. and Heine, B. (eds.), Approaches to grammaticalization*, pp.189–218, volume I. Oxford: Blackwell Publishing.

Traugott, Elizabeth Closs and Graeme Trousdale. 2013. *Constructionalization and constructional changes*, volume 6. Oxford University Press.
Trousdale, Graeme. 2014. On the relationship between grammaticalization and constructionalization. *Folia Linguistica*, 48(2):557–578.
Vetters, Carl and Audrey Lière. 2009. Quand une périphrase devient temps verbal: le cas d'aller+ infinitif. *Faits de langues*, 33(1):27–36.

Appendix 1. Texts included in the corpus and their corpora of origin

English translation corpus

corpus of origin	identifier	author	title	pub. year	decade	genre	source text identifier	reference text identifier
EEBO-TCP	ET01	Greene, Robert, 1558-1592.	Gvvydonius The carde of fancie wherein the folly of those carpet knights is decyphered, which guyding their course by the compasse of Cupid, either dash their ship against most daungerous rocks, or els attaine the hauen with paine and perill. Wherein also is described in the person of Gwydonius, a cruell combat betvveene nature and necessitie. By Robert Greene Master of Arte, in Cambridge.	1584	1580-1589	literature	EFS01	ER01
EEBO-TCP	ET02	Munday, Anthony	The queen of Nauarres tales Containing, verie pleasant discourses of fortunate louers. Now newly translated out of French into English.	1597	1590-1599	literature	EFS02	ER02
EEBO-TCP	ET03	Munday, Anthony	The defence of contraries Paradoxes against common opinion, debated in forme of declamations in place of publike censure: only to exercise yong wittes in difficult matters. Wherein is no offence to Gods honour, the estate of princes, or priuate mens honest actions: but pleasant recreation to beguile the iniquity of time. Translated out of French by A.M. one of the messengers of her Maiesties Chamber.	1593	1590-1599	philosophy /morality	EFS03	ER03
EEBO-TCP	ET04	Kyd, Thomas	Pompey the Great, his faire Corneliaes tragedie effected by her father and husbandes downe-cast, death, and fortune. Written in French, by that excellent poet Ro : Garnier; and translated into English by Thomas Kid.	1595	1590-1599	drama	EFS04	ER04
EEBO-TCP	ET05	Lennard, Samson	Of wisdome three bookes written in French by Peter Charro[n] Doctr of Lawe in Paris. Translated by Samson Lennard	1608	1600-1609	philosophy /morality	EFS05	ER05
EEBO-TCP	ET06	R.D.	The mirrour of mirth and pleasant conceits containing many proper and pleasaunt inuentions, for the recreation and delight of many, and to the hurt and hinderance of none / framed in French by that worshipfull and learned gentleman, Bonaduenture de Periers ... and Englished by R.D.	1583	1580-1589	literature	EFS06	ER06
EEBO-TCP	ET07	Wilcox, Thomas	Three propositions or speeches, which that excellent man M. Iohn Caluin, one of the pastors of the Church of God in Geneua had there To which also is added, an exposition vpon that parte of the catechisme, which is appointed for the three and fortieth Sunday in number. Translated into Englishe, by T.VV.	1580	1580-1589	religious	EFS07	ER07

EEBO-TCP	ET08	Arthur Hall Esquire	Ten books of Homers Iliades, translated out of French, by Arthur Hall Esquire	1581	1580-1589	poetry	EFS08	ER08
EEBO-TCP	ET09	Grimeston, Edward	Admirable and memorable histories containing the wonders of our time. Collected into French out of the best authors. By I. [sic] Goulart. And out of French into English. By Ed. Grimeston. The contents of this booke followe the authors aduertisement to the reader	1607	1600-1609	history	EFS09	ER09
EEBO-TCP	ET10	Aggas, Edward	The politicke and militarie discourses of the Lord de La Nouue VVhereunto are adioyned certaine obseruations of the same author, of things happened during the three late ciuill warres of France. With a true declaration of manie particulars touching the same. All faithfully translated out of the French by E.A.	1588	1580-1589	history	EFS10	ER10
EEBO-TCP	ET11	Munday, Anthony	The honorable, pleasant and rare conceited historie of Palmendos Sonne to the famous and fortunate Prince Palmerin d'Oliua, Emperour of Constantinople and the queene of Tharsus. Translated out of French by A.M. one of the messengers of her Maiesties chamber.	1589	1580-1589	literature	EFS11	ER11
EEBO-TCP	ET12	Church, Rooke	The historie of the troubles of Hungarie containing the pitifull losse and ruine of that kingdome, and the warres happened there, in that time, betweene the Christians and Turkes. By Mart. Fumée Lord of Genillé, Knight of the Kings order. Newly translated out of French into English, by R.C. Gentleman.	1600	1600-1609	history	EFS12	ER12
EEBO-TCP	ET13	Erondelle, P.	Noua Francia : or The description of that part of Nevv France, which is one continent with Virginia Described in the three late voyages and plantation made by Monsieur de Monts, Monsieur du Pont-Graué, and Monsieur de Poutrincourt, into the countries called by the Frenchmen La Cadie, lying to the southwest of Cape Breton. Together with an excellent seuerall treatie of all the commodities of the said countries, and maners of the naturall inhabitants of the same. Translated out of French into English by P.E.	1609	1600-1609	history	EFS13	ER13
EEBO-TCP	ET14	anon10	Fovvre bookes, of the institution, vse and doctrine of the holy sacrament of the Eucharist in the old Church As likevvise, hovv, vvhen, and by what degrees the masse is brought in, in place thereof. By my Lord Philip of Mornai, Lord of Plessis-Marli; councellor to the King in his councell of estate, captaine of fiftie men at armes in the Kings paie, gouernour of his towne and castle of Samur, ouerseer of his house and crowne of Nauarre.	1600	1600-1609	religious	EFS14	ER14

EEBO-TCP	ET15	Hakluyt, Richard, 1552?-1616.	A notable historie containing foure voyages made by certayne French captaynes vnto Florida vvherein the great riches and fruitefulnes of the countrey with the maners of the people hitherto concealed are brought to light, written all, sauing the last, by Monsieur Laudonniere, who remained there himselfe as the French Kings lieuetenant a yere and a quarter: newly translated out of French into English by R.H. In the end is added a large table for the better finding out the principall matters contayned in this worke.	1587	1580-1589	travel	EFS15	ER15
EEBO-TCP	ET16	Watson, William, 1559-1603.	The Iesuites catechisme. Or Examination of their doctrine. Published in French this present yeere 1602. and nowe translated into English. VVith a table at the end, of all the maine poynts that are disputed and handled therein	1602	1600-1609	religious	EFS16	ER16
EEBO-TCP	ET17	Hunton, Anthony	A worthy treatise of the eyes contayning the knowledge and cure of one hundred and thirtene diseases, incident vnto them: first gathered & written in French, by Iacques Guillemeau, chyrurgion to the French King, and now translated into English, togeather with a profitable treatise of the scorbie; & another of the cancer by A.H. Also next to the treatise of the eies is adoiyned a work touching the preseruation of the sight, set forth by VV. Bailey. D. of Phisick	1587	1580-1589	medical	EFS17	ER17
EEBO-TCP	ET18	Aggas, Edward	A necessary discourse concerning the right which the house of Guyze pretendeth to the crowne of France. Faithfully translated out of the French	1586	1580-1589	history	EFS18	ER18
EEBO-TCP	ET19	Munday, Anthony	The famous, pleasant, and variable historie, of Palladine of England Discoursing of honorable aduentures, of knightly deedes of armes and chiualrie: enterlaced likewise with the loue of sundrie noble personages, as time and affection limited their desires. ... Translated out of French by A.M. one of the messengers of her Maiesties Chamber.	1588	1580-1589	literature	EFS19	ER19
EEBO-TCP	ET20	Surphlet, Richard	A discourse of the preseruation of the sight : of melancholike diseases; of rheumes, and of old age. Composed by M. Andreas Laurentius, ordinarie phisition to the King, and publike professor of phisicke in the Vniuersitie of Mompelier. Translated out of French into English, according to the last edition, by Richard Surphlet, practitioner in phisicke	1599	1590-1599	medical	EFS20	ER20
EMMA	LT01	Behn, Aphra, 1640-1689.	A discovery of new worlds from the French, made English by A. Behn	1688	1680-1689	literature	LFS01	LR01
EMMA	LT02	Behn, Aphra, 1640-1689.	The history of oracles, and the cheats of the pagan priests in two parts / made English.	1688	1680-1689	religious	LFS01	LR02

EMMA	LT03	Dennis, John	Letters upon several occasions written by and between Mr. Dryden, Mr. Wycherly, Mr. ----, Mr. Congreve, and Mr. Dennis, published by Mr. Dennis with a new translation of select letters of Monsieur Voiture.	1696	1690-1699	literature	LFS03	LR03
EEBO-TCP	LT04	Taylor, Thomas	A voyage to the world of Cartesius written originally in French, translated into English by T. Taylor, of Magdalen Colledge in Oxford.	1694	1690-1699	philosophy /morality	LFS04	LR04
EEBO-TCP	LT05	Soames, William; Dryden, John	The art of poetry written in French by the Sieur de Boileau ; made English.	1683	1680-1689	poetry	LFS05	LR05
EEBO-TCP	LT06	anon01	A discourse on the history of the whole world dedicated to His Royal Highness, the Dauphin, and explicating the continuance of religion with the changes of states and empires, from the creation till the reign of Charles the Great / written originally in French by James Benigne Bossuet ... ; faithfully Englished.	1686	1680-1689	history	LFS06	LR06
EEBO-TCP	LT07	Stanhope, George, 1660-1728.	Of wisdom three books / written originally in French by the Sieur de Charron ; with an account of the author, made English by George Stanhope ...	1697	1690-1699	religious	LFS07	LR07
EEBO-TCP	LT08	Lovell, Archibald	The comical history of the states and empires of the worlds of the moon and sun written in French by Cyrano Bergerac ; and newly Englished by A. Lovell ...	1687	1680-1689	literature	LFS08 & LFS09	LR08
EEBO-TCP	LT10	anon02	Lucian's ghost : or, Dialogues between the dead, wandering in the Elyzian shades. Being certain satyrical remarques upon the vain ostentatious humours of several learned and philosophical men and women, as well ancient as modern. Composed first in French, and now paraphras'd into English, by a person of quality.	1684	1680-1689	literature	LFS10	LR10
EEBO-TCP	LT11	anon03	Logic, or, The art of thinking in which, besides the common, are contain'd many excellent new rules, very profitable for directing of reason and acquiring of judgment in things as well relating to the instruction of for the excellency of the matter printed many times in French and Latin, and now for publick good translated into English by several hands.	1685	1680-1689	philosophy /morality	LFS11	LR11
EEBO-TCP	LT12	anon04	The whole art of the stage containing not only the rules of the drammatick art, but many curious observations about it, which may be of great use to the authors, actors, and spectators of plays : together with much critical learning about the stage and plays of the antients / written in French by the command of Cardinal Richelieu by Monsieur Hedelin, Abbot of Aubignac, and now made English.	1684	1680-1689	drama	LFS12	LR12

EEBO-TCP	LT13	Lussan, Henry	A vindication of the truth of Christian religion against the objections of all modern opposers written in French by James Abbadie ... ; render'd into English by H.L.	1694	1690-1699	religious	LFS13	LR13
EEBO-TCP	LT14	S.R.	The life of Monsieur Des Cartes containing the history of his philosophy and works : as also the most remarkable things that befell him during the whole course of his life / translated from the French by S.R.	1693	1690-1699	philosophy /morality	LFS14	LR14
EEBO-TCP	LT15	anon05	The falshood of human virtue a moral essay / done out of French.	1691	1690-1699	philosophy /morality	LFS15	LR15
EEBO-TCP	LT16	anon06	The maxims of the saints explained, concerning the interiour life by the Lord Arch-bishop of Cambray &c. ; to which are added, Thirty-four articles by the Lord Arch-Bishop of Paris, the Bishops of Meaux and Chartres, (that occasioned this book), also their declaration upon it ; together with the French-King's and the Arch-Bishop of Cambray's letters to the Pope upon the same subject.	1698	1690-1699	history	LFS16	LR16
EEBO-TCP	LT17	anon07	The adventures of Telemachus, the son of Ulysses translated from the French.	1699	1690-1699	literature	LFS17	LR17
EEBO-TCP	LT18	anon08	Moral maxims and reflections in four parts / written in French by the Duke of Rochefoucault ; now made English.	1694	1690-1699	philosophy /morality	LFS18	LR18
EMMA	LT19	Dryden, John, 1631-1700.	The life of St. Francis Xavier, of the Society of Jesus, apostle of the Indies, and of Japan written in French by Father Dominick Bohours, of the same Society ; translated into English by Mr. Dryden.	1688	1680-1689	religious	LFS19	LR19
EMMA	LT20	Defoe, Daniel, 1661-1731.	An account of Monsieur De Quesne's late expedition at Chio together with the negotiation of Monsieur Guilleragues, the French ambassadour at the port / in a letter written by an officer of the Grand Vizir's to a pacha ; translated into English.	1683	1680-1689	history	LFS20	LR20

EEBO-TCP	LT21	Lovell, Archibald	The critical history of the religions and customs of the eastern nations written in French by the learned Father Simon ; and now done into English, by A. Lovell ...	1685	1680-1689	religious	LFS21	LR21

English reference corpus

corpus of origin	**identifier**	**author**	**title**	**pub. year**	**decade**	**genre**	**corresponding source text identifier**	**corresponding translation identifier**
EEBO-TCP	ER01	Greene, Robert, 1558-1592.	Morando the tritameron of loue wherein certaine pleasaunt conceites, vttered by diuers woorthy personages, are perfectly dyscoursed, and three doubtfull questyons of loue, most pithely and pleasauntly discussed: shewing to the wyse howe to vse loue, and to the fonde, howe to eschew lust: and yeelding to all both pleasure and profitt. By Robert Greene, Maister of Artes in Cambridge.	1584	1580-1589	literature	EFS01	ET01
EEBO-TCP	ER02	Middleton, Christopher, 1560?-1628.	The famous historie of Chinon of England with his strange aduentures for the loue of Celestina daughter to Lewis King of Fraunce. VVith the worthy atchiuement of Sir Lancelot du Lake, and Sir Tristram du Lions for fair Laura, daughter to Cador Earle of Cornewall, beeing all knights of King Arthurs round table. By Chr. Middleton.	1597	1590-1599	literature	EFS02	ET02

EEBO-TCP	ER03	Rich, Barnabe, 1540?-1617.	A martial conference pleasantly discoursed betweene two souldiers, the one Captaine Skil, trained vp in the French and Low Country seruices, the other Captaine Pill, only practised in Finsburie fields in the modern warres of the renowmed Duke of Shordich and the mightie Prince Arthur / newly translated out of Essex into English by Barnabe Rich ...	1598	1590-1599	philosophy /morality	EFS03	ET03
EEBO-TCP	ER04	Kyd, Thomas	The Spanish tragedie containing the lamentable end of Don Horatio, and Bel-imperia: with the pittifull death of olde Hieronimo.	1592	1590-1599	drama	EFS04	ET04
EEBO-TCP	ER05	Vaughan, William, 1577-1641.	The golden-groue moralized in three bookes: a worke very necessary for all such, as would know how to gouerne themselues, their houses, or their countrey. Made by W. Vaughan, Master of Artes, and student in the ciuill law,	1600	1600-1609	philosophy /morality	EFS05	ET05
EEBO-TCP	ER06	Melbancke, Brian.	Philotimus. The warre betwixt nature and fortune. Compiled by Brian Melbancke student in Graies Inne	1583	1580-1589	literature	EFS06	ET06
EEBO-TCP	ER07	Parsons, Robert, 1546-1610.	A brief discours contayning certayne reasons why Catholiques refuse to goe to church. Written by a learned and vertuous man, to a friend of his in England. And dedicated by I.H. to the Queenes most excellent Maiestie	1580	1580-1589	religious	EFS07	ET07
EEBO-TCP	ER08	Howell, Thomas, fl. 1568-1581.	H. His deuises, for his owne exercise, and his friends pleasure	1581	1580-1589	religious	EFS08	ET08
EEBO-TCP	ER09	Ayscu, Edward.	A historie contayning the vvarres, treaties, marriages, and other occurrents betweene England and Scotland from King William the Conqueror, vntill the happy vnion of them both in our gratious King Iames. With a briefe declaration of the first inhabitants of this island : and what seuerall nations haue sithence settled them-selues therein one after an other :	1607	1600-1609	history	EFS09	ET09
EEBO-TCP	ER10	Marten, Anthony, d. 1597.	An exhortation, to stirre vp the mindes of all her Maiesties faithfull subiects, to defend their countrey in this dangerous time, from the inuasion of enemies. Faithfullie and zealouslie compiled by Anthonie Marten, sewer of her Maiesties most honorable chamber	1588	1580-1589	history	EFS10	ET10
EEBO-TCP	ER11	Munday, Anthony	A discouerie of Edmund Campion, and his confederates, their most horrible and traiterous practises, against her Maiesties most royall person and the realme Wherein may be seene, how thorowe the whole course of their araignement: they were notably conuicted of euery cause. VVhereto is added, the execution of Edmund Campion, Raphe Sherwin, and Alexander Brian, executed at Tiborne the 1. of December. Published by A.M. sometime the Popes scholler, allowed in the seminarie at Roome amongst them: a discourse needefull to be read of euery man, to beware how they deale with such secret seducers. Seene, and allowed.	1580	1580-1589	literature	EFS11	ET11

EEBO-TCP	ER12	Danett, Thomas, fl. 1566-1601.	A continuation of the historie of France from the death of Charles the eight where Comines endeth, till the death of Henry the second. Collected by Thomas Danett Gentleman.	1600	1600-1609	history	EFS12	ET12
EEBO-TCP	ER13	Daniel, Samuel, 1562-1619.	The ciuile wars betweene the howses of Lancaster and Yorke corrected and continued by Samuel Daniel one of the groomes of hir Maiesties most honorable Priuie Chamber	1609	1600-1609	history	EFS13	ET13
EEBO-TCP	ER14	Johnson, Francis, 1562-1618.	An answer to Maister H. Iacob his defence of the churches and minstery of England. By Francis Iohnson an exile of Iesus Christ	1600	1600-1609	religious	EFS14	ET14
EEBO-TCP	ER15	Hakluyt, Richard, 1552?-1616.	Diuers voyages touching the discouerie of America, and the ilands adiacent vnto the same made first of all by our Englishmen, and afterward by the Frenchmen and Britons : and certaine notes of aduertisements for obseruations, necessarie for such as shall heereafter make the like attempt, with two mappes annexed heereunto for the plainer vnderstanding of the whole matter.	1582	1580-1589	travel	EFS15	ET15
EEBO-TCP	ER16	Watson, William, 1559-1603.	A decacordon of ten quodlibeticall questions concerning religion and state wherein the authour framing himfelfe [sic] a quilibet to euery quodlibet, decides an hundred crosse interrogatorie doubts, about the generall contentions betwixt the seminarie priests and Iesuits at this present.	1602	1600-1609	religious	EFS16	ET16
EEBO-TCP	ER17	Baley, Walter, 1529-1592.	A discourse of the medicine called mithridatium declaring the firste beginninge, the temperament, the noble vertues, and the true vse of the same: compiled rather for those which are to vse it, then for the learned.	1585	1580-1589	medical	EFS17	ET17
EEBO-TCP	ER18	Colville, John, 1542?-1605.	Ane declaratioun of the iust and necessar causis, moving us of the nobillitie of Scotland & vthers ye Kings Maiesteis faithful subiectis to repair to his Hienes presence and to remane with him for resisting of the present daingeris appearing to Goddis trew religion and professours thairof & to his Hienes awin person estait & croun & his faithful subiectis that hes constantly continuit in his obedience, &c to seik redres & reformatioun of the abuse and confusioun of the commoun wealth remouing fro[m] his Maiestie the cheif authoris thairof quhil the treuth of the famin may be made manifest to his Hienes estaits that with common consent redres & remeid may be prouidid.	1582	1580-1589	history	EFS18	ET18
EEBO-TCP	ER19	Munday, Anthony	A courtly controuersie, betweene looue and learning Pleasauntlie passed in disputation, betweene a ladie and a gentleman of Scienna. Wherein is no offence offered to the vertuous nor any ill motion to delight the vicious.	1581	1580-1589	literature	EFS19	ET19

EEBO-TCP	ER20	Levens, Peter, fl. 1587.	A right profitable booke for all diseases called, The pathway to health. Wherein are to be founde most excellent & approued medicines of great vertue : as also notable potions and drinks, and for the distilling of diuers precious waters, and making of oyles, and other comfortable receits for the health of the body, neuer before imprinted. First gathered by Peter Leuens, master of art of Oxford, and student in phisicke and surgery : and now newly corrected and augmented.	1596	1590-1599	medical	EFS20	ET20
EEBO-TCP	LR01	Behn, Aphra, 1640-1689.	The lucky mistake a new novel / written by Mrs. A. Behn	1689	1680-1689	literature	LFS01	LT01
EEBO-TCP	LR02	Behn, Aphra, 1640-1689.	The fair jilt, or, The history of Prince Tarquin and Miranda written by Mrs. A. Behn.	1688	1680-1689	literature	LFS01	LT02
EMMA	LR03	Dennis, John	Remarks on a book entituled Prince Arthur, an heroick poem with some general critical observations and several new remarks upon Virgil / by Mr. Dennis.	1696	1690-1699	literature	LFS03	LT03
EEBO-TCP	LR04	Pomfret, Samuel, d. 1722.	A sermon preach'd to young people January the first, 1697. And now publish'd at their request. By Samuel Pomfret.	1698	1690-1699	philosophy /morality	LFS04	LT04
EEBO-TCP	LR05	Heynes, Thomas.	The triumphs of royalty in the person of King Charles II a poem / by Thomas Heynes, Gent.	1683	1680-1689	poetry	LFS05	LT05
EEBO-TCP	LR06	Joyner, William, 1622-1706.	Some observations upon the life of Reginaldus Polus Cardinal of the royal bloud of England sent in a pacquet out of Wales, by G.L. gentleman, and servant to the late Majesty of Henrietta Maria of Bourbon, mother to the present King.	1686	1680-1689	history	LFS06	LT06
EEBO-TCP	LR07	Stanhope, George, 1660-1728.	A sermon preach'd in St. Paul's Cathedral at the annual meeting of the Sons of the Clergy, Tuesday, December vii, 1697 by George Stanhope ... : to which is annexed a faithful account of the charities distributed by that corporation for fourteen years last past.	1698	1690-1699	religious	LFS07	LT07
EEBO-TCP	LR08	Gibbs, Richard, fl. 1681-1687.	The new disorders of love. A gallant novel. Written by Richard Gibbs, of Norwich, philo. medici.	1687	1680-1689	literature	LFS08 & LFS09	LT08
EEBO-TCP	LR10	Charleton, Walter, 1619-1707.	The Cimmerian matron to which is added The mysteries and miracles of love / by P.M. gent.	1684	1680-1689	literature	LFS10	LT10

EEBO-TCP	LR11	Grenville, Denis, 1637-1703.	Counsel and directions divine and moral in plain and familiar letters of advice from a divine of the Church of England, to a young gentleman, his nephew, soon after his admission into a college in Oxford.	1685	1680-1689	philosophy /morality	LFS11	LT11
EEBO-TCP	LR12	Shakespeare, William, 1564-1616.	Julius Cæsar a tragedy : as it is now acted at the Theatre Royal / written by William Shakespeare.	1684	1680-1689	drama	LFS12	LT12
EEBO-TCP	LR13	Locke, John, 1632-1704.	Reason and religion in some useful reflections on the most eminent hypotheses concerning the first principles, and nature of things : with advice suitable to the subject, and seasonable for these times.	1694	1690-1699	religious	LFS13	LT13
EMMA	LR14	Penn, William, 1644-1718.	Some fruits of solitude in reflections and maxims relating to the conduct of human life. Licens'd, May 24. 1693.	1693	1690-1699	philosophy /morality	LFS14	LT14
EEBO-TCP	LR15	Nourse, Timothy, d. 1699.	A discourse of natural and reveal'd religion in several essays, or, The light of nature a guide to divine truth.	1691	1690-1699	philosophy /morality	LFS15	LT15
EEBO-TCP	LR16	Trenchard, John, 1662-1723.	A short history of standing armies in England	1698	1690-1699	history	LFS16	LT16
EEBO-TCP	LR17	Winstanley, William, 1628?-1698.	The Essex champion, or, The famous history of Sir Billy of Billerecay and his Squire Ricardo shewing the renowned atchievements and valiant acts of the invincible knight Sir Billy of Billerecay with his love to his Lady Dulcinia, as also, the comical exploits and amorous adventures of Squire Ricardo performed by his invisible ring, their tilts and tournaments in honour of ladies, their battles with gyants and monsters, their punishing of inchanters and negromancers, pedlars and poppit-players / being that excellent piece of knight-errantry written by William Winstandly.	1699	1690-1699	literature	LFS17	LT17
EEBO-TCP	LR18	Allestree, Richard, 1619-1681.	The government of the thoughts a prefatory discourse to The government of the tongue / by the author of The whole duty of man.	1694	1690-1699	philosophy /morality	LFS18	LT18

EMMA	LR19	Dryden, John, 1631-1700.	A defence of the papers written by the late king of blessed memory, and Duchess of York, against the answer made to them	1686	1680-1689	religious	LFS19	LT19
EMMA	LR20	Defoe, Daniel, 1661-1731.	Reflections upon the late great revolution written by a lay-hand in the country for the satisfaction of some neighbours.	1689	1680-1689	history	LFS20	LT20
EEBO-TCP	LR21	Mackenzie, George, Sir, 1636-1691.	A defence of the antiquity of the royal line of Scotland with a true account when the Scots were govern'd by kings in the isle of Britain / by Sir George Mackenzie ...	1685	1680-1689	religious	LFS21	LT21

French source text corpus

corpus of origin	**identifier**	**author**	**title**	**pub. year**	**genre**	**corresponding translation identifier**	**corresponding reference text identifier**
Frantext	EFS01	Labé, Louise	Débat de folie et d'amour	1555	literature	ET01	ER01
Frantext	EFS02	Navarre, Marguerite	Histoires des amans fortunez, commonly known as L'Heptaméron.	1559	literature	ET02	ER02
Frantext	EFS03	Estienne, Charles	Paradoxes, ce sont propos contre la commune opinion	1561	philosophy /morality	ET03	ER03
Frantext	EFS04	Garnier, Robert	Cornélie	1585	drama	ET04	ER04
Frantext	EFS05	Charron, Pierre	De la sagesse	1601	philosophy /morality	ET05	ER05

Frantext	EFS06	Des Périers, Bonaventure	Nouvelles récréations et joyeux devis (1558)	1558	literature	ET06	ER06
Google Books	EFS07	Calvin, Jean	Catéchisme de l'Eglise de Genève	1576	religious	ET07	ER07
Google Books	EFS08	Salel, Hugues	L'Iliade (Homère) . French version/original: L'Iliade d'Homère traduit du grec en françois, par Hugues Salel	1574	poetry	ET08	ER08
Google Books	EFS09	Goulart, Simon	Histoire admirables et memorables de nostre temps	1604	history	ET09	ER09
Google Books	EFS10	La Noue, François de	Discours politiques et militaires du Seigneur de la Noue.	1587	history	ET10	ER10
Google Books	EFS11	Vernassal, François de	L'Histoire de Primaleon de Grece continuant celle de Palmerin d'Olive (François de Vernassal)	1550	literature	ET11	ER11
Google Books	EFS12	Fumée, Martin	Histoire des troubles de Hongrie.	1595	history	ET12	ER12
Google Books	EFS13	Lescarbot, Marc	Histoire de la Nouvelle France (Lescarbot, Marc)	1609	history	ET13	ER13
Google Books	EFS14	Mornay, Philippe de, seigneur du Plessis-Marly, 1549-1623.	De l'institution, usage, et doctrine du sainct sacrement de l'Eucharistie, en l'eglise ancienne. Mornay, Philippe	1598	religious	ET14	ER14
Google Books	EFS15	Laudonnière, René Goulaine de.	L'histoire notable de la Floride	1587	travel	ET15	ER15
Google Books	EFS16	Pasquier, Etienne, 1529-1615.	Le catechisme des Jesuites.	1502	religious	ET16	ER16

Google Books	EFS17	Guilleme au, Jacques	Traité des maladies de l'oeil	1585	medical	ET17	ER17
Google Books	EFS18	Mornay, Philippe de, seigneur du Plessis-Marly, 1549-1623.	Discours sur le droit prétendu par ceux de Guise sur la couronne de France	1583	history	ET18	ER18
Google Books	EFS19	Colet, Claude, 16th cent.	Histoire palladienne	1562	literature	ET19	ER19
Google Books	EFS20	Du Laurens, André, 1558-1609.	Discours de la conservation de la veüe: des maladies melancholiques: des catarrhes: et de la vieilless	1597	medical	ET20	ER20
Frantext	LFS01	Fontenell e, M.	Entretiens sur la pluralité des mondes	1686	literature	LT01	LR01 & LR(
Frantext	LFS03	Voiture, Vincent	Lettres	1648	literature	LT03	LR03
Frantext	LFS04	Daniel, Gabriel	Voiage du monde de Descartes	1690	philosophy /morality	LT04	LR04
Frantext	LFS05	Boileau Despréau x, Nicolas, 1636-1711.	L'art poétique	1674	poetry	LT05	LR05
Frantext	LFS06	Bossuet, Jacques Bénigne, 1627-1704.	Discours sur l'histoire universelle à Monseigneur le Dauphin : pour expliquer la suite de la religion et les changemens des empires	1681	history	LT06	LR06

Frantext	LFS07	Charron, Pierre	De la sagesse	1601	religious	LT07	LR07
Frantext	LFS08	Cyrano de Bergerac	Les Estats et empires de la lune	1655	literature	LT08	LR08
Frantext	LFS09	Cyrano de Bergerac	Les Estats et empires du soleil	1655	literature	LT09	LR09
Frantext	LFS10	Fontenelle, M.	Digression sur les anciens et les modernes	1688	literature	LT10	LR10
Frantext	LFS11	Arnauld, Antoine; Nicole, Pierre	La logique, ou l'art de penser	1662	philosophy /morality	LT11	LR11
Frantext	LFS12	Aubignac, François-Hédelin, abbé d', 1604-1676.	La pratique du théatre.	1657	drama	LT12	LR12
Frantext	LFS13	Abbadie, Jacques, 1654-1727.	Traité de la verité de la religion chrétienne.	1684	religious	LT13	LR13
Frantext	LFS14	Baillet, Adrien, 1649-1706.	La Vie de Monsieur Descartes	1691	philosophy /morality	LT14	LR14
Frantext	LFS15	Esprit, Mr. 1611-1678.	La fausseté des vertus humaines	1678	philosophy /morality	LT15	LR15

Frantext	LFS16	Fénelon, François de Salignac de La Mothe-, 1651-1715.	Explication des maximes des saints	1697	history	LT16	LR16
Frantext	LFS17	Fénelon, François de Salignac de La Mothe-, 1651-1715.	Les aventures de Télémaque	1699	literature	LT17	LR17
Frantext	LFS18	La Rochefoucauld, François, duc de, 1613-1680.	Maximes	1665	philosophy /morality	LT18	LR18
Google Books	LFS19	Bouhours, Dominique, 1628-1702.	La vie de Saint François Xavier de la Campagnie de Jesus, apostre des Indes et du Japon	1682	religious	LT19	LR19
Google Books	LFS20	Officer of the Grand Vizir.	Substance d'une lettre écrite par un Officier du GrandVizir un pacha, touchant l'expedition de Monsr du Quesne à Chio et la négotiation de Monsr de Guilleragues avec la Port.	1683	history	LT20	LR20
Google Books	LFS21	Simon, Richard, 1638-1712.	Histoire critique de la creance et de coutumes des nations du Levant	1684	religious	LT21	LR21

Appendix 2. Used Regex queries (Regular Expressions)
Claes Pauline. Master of Linguistics (Digital Text Analysis). Academic Year 2021-2022. Master's Thesis.

This files contains all Regular Expressions queries used for retrieving concordances. All regular expressions used were adopted from existing regular expressions previously stored in CosyCat (Arévalo and Petré 2017) by Prof. Petré in the contexts of one of his MA courses and will be added as Appendix 2.

Progressive [BE Ving]+[GO to V]
For the texts included in the EMMA corpus, the following Regex was provided by Cosycat (Arévalo and Petré 2017):

".*'[sſmr]e?|am|ar[te]?|i[sſ]|wa[sſ]t?|wer[et]?|be+n?e?|be+[yi]nge?" [!word="\.|;|,|from|in|of|at|after|like|for|my|a|-"]{0,2} "a"? "-"? "a?goe?ing|_oing|g_ing" [!word="to|\."]{0,5} "to" [!word="\.|;|,|\!|\?|:"]

This Regex searches for all instances of be going to + INF that meet the following conditions:
- any form of \textit{to be}, including contracted forms and historical spellings (am, 'm, are, 're, is, 's, art, wert, ...);
- typical forms and spellings of going (going, a going, agoing, a-going);
- allowing for maximally two words between the copula and the form of going;
- allowing for maximally five words between the form of going and to.

For the texts collected from EEBO, a regular expression was used that searches for all progressive instances of the verb go:

"a?goe?ing|_oing|g_ing"

Non-progressive [GO to V]
Both the texts collected from EMMA as from EEBO were queried for non-progressive instances of [GO to V] using the following regular expression provided by Cosycat (Arévalo and Petré 2017):

"(go+e?[ts]?h?e?|gaes?|wente?s?t?|vvente?s?t?)" "to"

AND:

(go+e?[tſ]?h?e?|gaeſ?|wente?ſ?t?|vvente?ſ?t?)''to' (including the historical long 's').