Anglicisms in translation

An analysis of English loans in Dutch original and translated cookbooks

Bonnie Dekker
b.dekker.2@umail.leidenuniv.nl
bonniedekker@chello.nl

Katinka Zeven (supervisor)
Lettie Dorst (second reader)

16 June, 2014

Table of Contents

List of tables, figures, and abbreviations

Abbreviations

NL-OR: the corpus containing original Dutch texts
NL-TR: the corpus containing Dutch texts that have been translated from English
EN-OR: the corpus containing the English source texts of the translations in NL-TR

Chapter 1: Introduction

1.1 Overview

This thesis compares the use of English borrowings, i.e. anglicisms, in Dutch original and translated cookbooks. The main purpose is to determine whether translators' tendency to explain and clarify causes them to produce translations that contain fewer anglicisms than similar original Dutch texts. The terms "loan" and "borrowing" can refer both to the process in which a speaker transfers an element from one language into another and to the result of that process; the exact definition of the word "anglicism" used for this thesis is explained in more detail in sections 2.3 and 3.3. This chapter will list the research questions, briefly outline the main theories that motivate these questions, and provide a short overview of the following chapters.

1.2 Theoretical background

The method used in this thesis is based on theory from the fields of corpus-based translation studies (providing a method of analysing translational text in comparison with non-translational text) and contact linguistics (providing information on the process and products of linguistic borrowing in general). Section 2.4 explains the notion that translational text is inherently different from non-translational text. Baker (1993) and Kruger (2002) identify a number of "translation universals", i.e. typical features of translated text that differentiate them from their source texts and from original texts written in the target language. Translators appear to be particularly inclined towards explicitation; translations tend to be more cohesively explicit and longer than their source texts (Blum-Kulka, 1986).

Previous studies into the use of borrowings in translated text as compared to original text have been performed by Frankenberg-Garcia (2005), Musacchio (2005), and Laviosa (2007). These studies focus on different language pairs (English-Portuguese and English-Italian), but their methods and findings may be generalisable to other language pairs as well. This thesis aims to examine the characteristics of borrowing in English to Dutch translation and to contrast these findings with the results found for the other languages.

1.3 Research questions

As discussed above, translations may be inherently different from non-translations, and the goal of this thesis is to compare the use of anglicisms in Dutch translational and non-translational text. More specifically, the research questions are:
  (1) Do cookbooks that have been translated from English into Dutch contain
       fewer anglicisms than those that were originally written in Dutch?

(2) Are there any differences between the anglicisms in translations and original texts in terms of type, function, and grammatical category?
(3) Do translations contain anglicisms that are more conventional than those in non-translational text?

1.4 Thesis overview

Chapter 2 provides an overview of theories and studies that are relevant to this thesis. This includes information on borrowing as a translation procedure, the possible motivations behind linguistic borrowing and the forms it can take, the characteristics of translational language compared to original language, the compilation and utilisation of corpora for translation studies, the prevalence of and attitudes towards anglicisms in the Netherlands, and studies that compare borrowings in translational and non-translational text for other languages. Chapter 3 describes the corpus selection process and the methods of classification and analysis. Chapter 4 presents the results of these methods and contrasts them to the literature described in chapter 2. Finally, chapter 5 sums up the relevant findings in order to answer the research questions.

Chapter 2: Literature

2.1 Introduction

This chapter summarises theory on the subject of linguistic borrowing both in translation and in general. Section 2.2 discusses perspectives on borrowing as a translation procedure, highlighting situations in which this method is considered appropriate and those in which it is better avoided. Section 2.3 discusses anglicisms in contexts beyond translation, including theories on identifying and classifying them. In order to introduce the notion of studying translated text as a phenomenon on its own that is different both from its source text and from non-translated text, section 2.4 discusses possible universal features of translation. To explain the methodology used for this thesis, section 2.5 introduces the field of corpus-based translation studies and discusses which types of corpus can be used for which purpose. Section 2.6 discusses two articles that illustrate the status and perception of anglicisms in the Netherlands. Section 2.7 summarises a number of studies that relate to the topic of this thesis in terms of their subject and/or method. Finally, section 2.8 summarises the theories that are most relevant to the research questions and discusses expectations as to the results based on the information gathered from the literature.

2.2 Borrowing as a translation procedure

"Borrowing" a word from the source text and inserting it directly into the target text may be the "simplest of all translation methods" (Vinay & Darbelnet, 2000, p. 85), but there are certainly situations where it seems appropriate or even necessary. For instance, Newmark explains that transference is customary for certain proper nouns such as the names of locations, people, and companies. He does advise to combine this method with another procedure into a translation couplet, for instance through the addition of an explanation of functional equivalent between brackets (Newmark, 1988, pp. 81-82).

　　　The decision to borrow depends on the text type, the intended readership, and their degree of competence in the source language. The more specialised the text and the more expert the readership, the more likely it is that the translator will need to transfer some terms from the source text, such as titles, cultural terms, and words that are used in a specific sense (Newmark, 1988, p. 100). This is particularly important if there is a chance that these expert readers will want to look for the term in other sources on the topic or consult the source text, as the inclusion of the source language word in the translation makes it easier for readers to recognise the concept elsewhere. In specialised contexts, every transferred term allows the reader to get closer to the sense of the original text. If the readership is likely to consist of people with varying degrees of competence in the source language, adding an explanation to the borrowing will ensure that all readers understand. The combination of the two

terms will signal to the reader that the relationship between the source and target terms is more complex than pure equivalence and will invite them to "envisage the gap mentally" (Newmark, 1988, p. 101).

In addition to mere semantic precision and recognisability of the source term, there may also be stylistic reasons that motivate the translator to borrow source text words. In novels, for instance, transferred words may provide "local colour" because the evoked image or sound of the term is attractive, while the same terms would be translated with a functional equivalent in other contexts. However, Newmark also warns against overuse of foreign words, noting that transference sometimes happens for "snob reasons" by translators who treat cultural terms as untranslatable because they are "posh" foreign words. Overall, he argues that it remains the translator's job to explain and make readers understand concepts from the source text, not to mystify them "by using vogue-words" (Newmark, 1988, p. 82).

2.3 Anglicisms in general

Motivations for the use of anglicisms

According to Haugen (1950, p. 212), borrowing occurs when a speaker attempts to reproduce patterns previously found in one language into another. In addition to the situations in which translators use borrowings, there are a variety of reasons to borrow that apply to all speakers of a language. The two broad categories into which loans are often divided are cultural borrowings (which have no equivalent in the native language) and core borrowings (for which a native equivalent already exists) (Myers-Scott, 2006). Cultural borrowings often enter a language along with new inventions and products (e.g. computers) and they are sometimes referred to as "necessary borrowings"—although borrowing is certainly not the only way for a language to acquire new words. Core borrowings—or "luxury loans" (Onysko & Winter-Froemel, 2011)—may be adopted for a variety of reasons.

Onysko groups the reasons why German speakers use anglicisms together into six motivations:
(1) semantic (e.g. for new products and inventions);
(2) stylistic (to avoid repetition of the same term);
(3) euphemistic (e.g. to avoid words that are taboo in the native language);
(4) emotive (i.e. because English sounds "modern, hip, and educated");
(5) social (to establish a sense of group identity); and
(6) brevity (for English words that are conveniently shorter than their native equivalents) (Onysko, 2004, pp. 62-63).

Onysko's division is similar to the one proposed by Galinsky (1967), who also mentions (a) variation of expression, (b) brevity, and (c) euphemism, in addition to four other motivations:
(d) to convey an American atmosphere or setting;
(e) for precision (e.g. due to different connotations);

    (f) metaphorical translations for the sake of vividness (i.e. loan translations such as *Wolkenkratzer* for *skyscraper*); and

    (g) for a comic touch or satire (Hilgendorf, 1996, pp. 5-8).

Borrowing may also occur as a way to avoid homonyms if a sound change makes two native words too similar (Haspelmath, 2009, p. 50).

      Clearly, there are many practical and stylistic purposes that motivate speakers to borrow words from another language. However, many of these could also be fulfilled using word formation processes within the speakers' native language. The fact that speakers choose borrowings over native neologisms can be attributed to the prestige of a dominant language (Haspelmath, 2009, pp. 46-49), in this case English.

Identifying anglicisms

For the analysis of a language's anglicisms, the exact definition of what constitutes an anglicism and the method used to recognise one will depend on the aim of the study. For the compilation of his *Dictionary of European Anglicisms*, Manfred Görlach selected words that were recognisably English in their form (orthographically, phonologically, and/or morphologically), but were accepted as items in the receptor language's vocabulary (Görlach, 2003, p. 1). This definition excludes words that have not been generally accepted by the speakers of the language as well as words that have been adapted so much that they no longer stand out as English to most speakers.

      The definition of the word *anglicisme* employed by the Genootschap Onze Taal, a society dedicated to the Dutch language, exemplifies a very different approach: it characterises *anglicismen* as loan translations from English that are generally considered to be incorrect and have often originated from "lazy translations". This definition includes lexical items as well as expressions that are the result of structural influence. Onze Taal's article explaining this concept acknowledges that speakers' view on the correctness of these anglicisms may change over time, but the definition also shows a degree of prescriptivism and it is followed by a list of anglicisms with their "acceptable" Dutch equivalents. ("Anglicismen", n.d.). Onze Taal's list of anglicisms that are currently considered unacceptable forms a useful tool to determine the degree of conventionalism of anglicisms in a corpus (see section 4.4 below), but it is too restrictive to be used on its own in a study that aims to analyse a variety of English borrowings in a corpus.

      Gottlieb suggests a broader definition of anglicisms; it includes any language feature that has either been adopted or adapted from English or has experienced a boost as a result of English influence. This description is intended to be all-inclusive and "cover the entire spectrum of present-day influence from English". It incorporates phenomena that would not appear in Görlach's dictionary, such as grammatical borrowing, new and ad hoc loans that have not become widely accepted, and native language features that have become more common due to English influence (Gottlieb, 2004, p. 44).

In order for a word to be classified under one of these definitions, it needs to be part of the following scenario: there must be a plausible situation of language contact, the word must be similar in shape and meaning to a word from the hypothetical source language, and there may not be any other plausible explanations for these similarities. Other explanations may be that the languages share a common ancestor through which they both acquired the word or that the borrowing process actually took place the other way around. The donor language can often be identified by examining its morphology (borrowings are usually morphologically analysable in one language but not the other), its phonology (the word may be phonologically integrated in only one of the languages), or its meaning (which may be more relevant to one of the two cultures) or by looking at the same word in sister languages (Haspelmath, 2009, p. 43-44). The main resource that was used to determine the etymology of the anglicisms discussed in this thesis is the online *Etymologiebank* (Van der Sijs, 2010).

For the purpose of this thesis, the definition of what constitutes an anglicism focuses mainly on lexical items without posing limitations on their degree of conventionalisation or acceptance. The decision to concentrate on lexical items is primarily a practical one, as they are simpler to identify than structural types of borrowing, and examining all lexical anglicisms in a text rather than only conventionalised loans seems like a more thorough way to analyse the authors' approach to borrowing. The process of defining and identifying borrowings within the corpora used for this thesis is explained more extensively in section 3.3.

Classifying anglicisms

Anglicisms may be subdivided into a wide variety of classes—Gottlieb's (2004) taxonomy includes fifteen categories, each further divided into several different types—but the types that are mentioned most often are loan words (which copy both meaning and phonemic shape, usually substituting native phonemes), hybrids (borrowings that are partly native and partly imported), loan translations or calques (in which the components of a foreign word are all replaced by native translations), and semantic loans (native words that expand their meaning to include the meaning of a foreign word). Haugen categorises these based on the criteria of importation and substitution, resulting in three main types:

   (1) loan words, which are the result of morphemic importation from the donor language but not substitution from the recipient language;

   (2) loan blends, which are subject to both morphemic importation and substitution of native elements; and

   (3) loan shifts, which show substitution of native elements but no importation of foreign morphemes.

In this categorisation, the previously mentioned hybrid would be considered a loan blend, and calques and semantic loans fall under loan shifts (Haugen, 1950, pp. 213-220).

Gottlieb employs a different classification for his typology of anglicisms in Danish, which is based on two main distinctions: first, items that are adopted or adapted into the recipient language on the one hand and items that are inspired or "numerically boosted" by phenomena from the English language on the other, and second, the distinction between microlanguage (i.e. the level of morphemes, phonemes, phraseology, etc.) and macrolanguage (i.e. the clause, sentence, or text level). This distinction leads him to divide anglicisms into three groups:

> (1) active anglicisms (sub-clause items that have been adopted or adapted from English, e.g. lexical borrowing, loan translations, and hybrids);
> (2) reactive anglicisms (sub-clause items that have been inspired or boosted by English models, e.g. semantic loans and orthographic loans); and
> (3) code-shifts (clause, sentence, and text items that have been adapted or adopted from English, e.g. sentence-shaped shifts and shift of full texts) (Gottlieb, 2004, pp. 44-48).

In addition to classifications based on the composition of the borrowing, loans have been sorted based on grammatical category in order to determine which categories are borrowed more often than others. Van Hout & Muysken (1994) cite several of these hierarchies of borrowability which suggest that nouns are the most susceptible to borrowing, followed first by adjectives and verbs and then by other parts of speech. In a later article, Muysken suggests that looking to develop a universal hierarchy may not be worthwhile, but he does list a number of specific hypothetical hierarchies, with the rightmost item being the most likely candidate for borrowing, e.g. for colours ("basic colours > peripheral colours"), numbers ("low numbers > high numbers"), and types of vocabulary ("core vocabulary > non-core vocabulary > animal and plant names > technical vocabulary") (Muysken, 2010, pp. 269-271).

Gottlieb also suggests a "hierarchy of success" that shows the various stages in the process of acceptance for anglicisms in Danish. At the bottom of this hierarchy are what Gottlieb calls peripheral anglicisms or non-accepted items. These are, in order of least to most likely to survive:

> (4) interfering items (such as mistranslations); and
> (3) implants (which still "sound" English and which are only accepted within certain user groups).

High on the "anglicism ladder of success" are the established anglicisms or accepted items:

> (2) naturalised items (which are identified as English loans and commonly accepted); and
> (1) integrated items (words that are not intuitively identified as English).

As these categories indicate, borrowings tend to go through a process of integration before becoming fully accepted, and many never make it to the top; "prospective anglicisms often die young" (Gottlieb, 2004, pp. 54-55).

2.4 Translational language: the third code

In order to analyse how exactly translators use anglicisms, it is necessary to examine translations both compared to their source texts and to original texts written in the same language. Frawley (1984) argues that the confrontation between the two languages during translation results in a communicative event that merits attention in its own right, i.e. the "third code" (Kruger, 2002, p. 80). This concept enables Frawley to quantify translations based on their degree of semiotic innovation, i.e. how much new knowledge they produce (Venuti, 2000, pp. 216). Previously, any way in which translations were "different" used to be seen as negative, "a sign of loss inherent in the translation process" (Tymoczko, 1998, p. 6), but moving beyond mere criticism and prescriptivism and examining the features that make translations unique can provide valuable insights into the translation process.

Translations, like all texts, are communicative events that take shape as a result of the goals and pressures of their own immediate context (Baker, 1996). Through an analysis of translations through corpora, Baker identifies the following universal features of translation:
   (1) explicitation;
   (2) disambiguation and simplification;
   (3) textual conventionality in translated novels;
   (4) avoidance of repetition present in the source text;
   (5) exaggeration of features of the target language; and
   (6) specific distribution of lexical items (Baker, 1993).
Kruger groups these features together into three, more general universals:
   (1) a tendency towards explicitation;
   (2) a tendency towards disambiguation; and
   (3) a tendency towards conventionalisation (Kruger, 2002, p. 81).

The notion of explicitation as a universal of translation is a prominent one. Blum-Kulka puts forth the explicitation hypothesis, which states that target texts are generally more cohesively explicit than their source texts, regardless of the characteristics of the two languages involved, because explicitation is an inherent process of translation (Blum-Kulka, 1986, p. 19). This hypothesis is supported by Frankenberg-Garcia's 2009 study which analysed explicitation in translations in terms of text length and found that target texts do tend to be longer than source texts (Frankenberg-Garcia, 2009a). These universal features of translation could influence translators' use of anglicisms, as well: an inclination towards explicitation or simplification may lead them to avoid borrowing and opt for an explanation or a hypernym instead.

In an attempt to formulate general laws of translation, Toury (1995) identifies two other norms. The first is the law of growing standardisation, which states that when no other conditions have been specified, textual relations from the source text tend to be omitted or modified to be more like the relations that are common in the target language. The second norm addresses influence in the opposite direction: the

law of interference states that features of the make-up of the source language tend to be transferred to the target text. Toury indicates the importance of the relationship between the two languages at play; tolerance of source language interference becomes greater if the source is a major language with a dominant, prestigious culture (Toury, 1995, pp. 267-279).

While the dominance of one language over the other will likely result in the translator transferring features from that language, the interplay between the two languages can also result in a kind of "levelling out" as translations tend to find a middle ground between two extremes. As a result, texts in a corpus of English translations are more similar to each other in terms of lexical density, mean sentence length and type/token ratio than texts in a comparable corpus of original English texts (Baker, 1996, p. 184). The two languages may also converge when it comes to borrowing; the loans cause foreign lexical patterns in translations that would not normally occur in the source or target language (Kruger, 2002, p. 80). Finally, the distinctive patterns that form in the translation compared to the source and target languages may also be a result of the translator's strategy, e.g. whether their intention is to foreignise or domesticise (Laviosa, 2002, p. 24).

2.5 Corpora and translation studies

Corpus-based translation studies

Corpus-based translation studies emerged in the 1990s as a combination of the fields of translation studies and corpus linguistics. The use of corpora has numerous benefits that facilitate research in this area. First of all, corpora allow users to extract data from large collections of texts that would be impractical to analyse manually and to use them for a variety of purposes including language learning, translation, and linguistic and cultural research. Corpora can be made available worldwide relatively easily, enabling and encouraging researchers to work together in team projects or replicate each other's research by investigating the same data. Moreover, corpora can be saved and expanded over time so that they can serve for extensive research as well as preservation of the data (Baker, 2007).

Despite the obvious benefits, this new technology also introduces a number of challenges. The fact that corpora provide so many opportunities to generate data and statistics makes it all the more important to remain focused on the main purpose of a research. Baker warns against a strong temptation to use statistics about translation to emphasise norms; too much focus on these norms cause users of corpora to label any translation that deviates from them "wrong". Instead, these norms should provide insight into universal features of translation and serve as a backdrop for the analysis of the more creative translation choices (Baker, 1996, p. 179).

For optimum results, the new technology of corpus linguistics should be used alongside traditional methods for translation studies, "not at the expense of human creativity and experience" (Baker, 2007). Tymoczko also advises users of corpora to

avoid "empty exercises" that emphasise quantification over substantive investigation, noting that the value of corpus-based translation does not lie in objectivity, but in the researcher's insightful interpretation of the data. The compilation of corpora, the design of experiments and the interpretation of data all depend on human judgment and intuition (Tymoczko, 1998, p. 3-8). Corpus users may enrich the data by considering socio-cultural issues and turning to information outside the corpus such as statements by authorities on the subject or the translators, authors, and publishers themselves.

      One of the drawbacks to working with corpora is the amount of time and money that goes into their creation. Compiling a large corpus often requires the work of a team of people with a range of expertises—in administration, linguistics, and computing, at the very least—and process of selecting, sampling, digitalising, and annotating texts, as well as requesting copyright permission (if the corpus is to be published online) can demand a lot of time (Baker, 2007, p. 52). Nevertheless, building a corpus for smaller projects—e.g. an ad hoc or "quick-and-dirty" corpus (Nesselhaulf, 2007, p. 298)—does not require quite such large investments.

Parallel and comparable corpora

      Corpus-based translation studies makes use of parallel corpora, which consist of source texts and their translations and allow the user to examine specific translation patterns, as well as comparable corpora, which consist of original texts in two or more languages and allow the user to compare patterns that occur naturally in each language. For comparable corpora, it is important to make sure the texts are similar in as many ways as possible within each language—e.g. the domain they cover, the variety of language, the length, and the range of authors and translators who produced them (Kruger, 2002, p. 87).

      In some cases, a bidirectional parallel corpus may also fulfil the purpose of a comparable corpus as it contains original texts from both languages. However, Zanettin points out that if the non-translational component of the corpus only consists of texts that have been translated (because they serve as source texts for the translational sub-corpus), then the corpus is not necessarily representative of all texts of that kind within the source language—just the texts that were chosen to be published abroad. The majority of texts produced in any given language are never translated, and perhaps the texts in the non-translational part of a comparable corpus share certain characteristics that are less common in the texts that fall outside the corpus. As Zanettin claims, "no language can be represented by a corpus which includes only texts that have been translated" (Zanettin, 2002, p. 330).

      Similarly, a corpus consisting of only original texts would not be representative of all written text production in a language—translations also form a part of that group. In order to be fully representative of the source language, then, the texts in a comparable corpus must be selected from the entire population of texts written in that language. For the analysis of translational text in comparison to non-

translational text, though, it is essential that the comparable component only consists of original texts. The exact characteristics of the corpora used for this thesis and the process of compilation and analysis will be described in the following chapter.

2.6 Anglicisms in Dutch: frequency, attitudes, and comprehension

Loan words in Dutch newspaper articles from 1994 and 2012

In a 2012 article, Van der Sijs responds to the general sentiment expressed by Dutch speakers (e.g. in letters to the editor) that the number of English borrowings in Dutch is growing at an alarming rate and at the expense of speakers' native language—some sources claim that 75% of Dutch vocabulary is derived from other languages. Van der Sijs investigates this issue by counting the number of loan words in samples from one recent (2012) and one older (1994) edition of the Dutch newspaper *NRC Handelsblad*. This analysis includes loan words from all languages (though she highlights anglicisms in particular) and excludes potential loan translations for which the etymology is uncertain.

Contrary to what seems to be popular opinion, the results of this study do not show a dramatic increase in the total percentage of loan words: borrowed words account for around 30% of the vocabulary (types) in each sample (and 16% of all tokens). For English, the results do show a small difference: out of all types, 2.3% of the 1994 sample and 3.7% of the 2012 sample are derived from English. However, Van der Sijs points out that this difference is not significant enough make any generalisations about the status of English loan words in general, particularly because the corpus is so small (11,314 words) and derived from only two newspapers. She also notes that while anglicisms are frequently used in advertising and TV, sometimes to the annoyance of viewers, these terms rarely last very long; English titles and taglines disappear along with the corresponding programmes and commercials (Van der Sijs, 2012).

The studies on anglicisms in translation described in section 2.8 are all based on languages other than Dutch. The article by Van der Sijs sheds some light on the presence and perception of anglicisms in the Netherlands. She notes that while the use of English in Dutch is increasing slightly, the new borrowings rarely survive very long. However, their short existence may still have a significant effect on speakers' perception. If these short-lived loans are always replaced with new borrowings, then the presence of English remains prominent. Judging by this article, Dutch speakers certainly seem to be very aware of (and sometimes irritated by) the existence of these anglicisms. Because of their salience, it seems important to include these transient borrowings in addition to the more established loans when analysing contemporary use of English in Dutch.

English in Dutch commercials

Like Van der Sijs, Gerritsen, Korzilius, Van Meurs, and Gijsbers (2000) observe that many publications in Dutch address the increasing pervasiveness of English, often in a negative manner, but they note that not much research has been done into the actual frequency of anglicisms in Dutch. Their study into the comprehension of and attitudes towards English in commercials on Dutch television shows that one third of the commercials they selected contained some form of English. The main reasons why companies advertise using English seem to be (1) to save costs by not having the text translated for each country where the product is marketed and (2) because, in the Netherlands, "everyone understands English anyway" (p. 18). This study, however, shows that these motivations are not necessarily valid: many of the viewers do not understand the meaning of the English segments, and if the misunderstanding affects their perception of the product negatively, the use of English in commercials may actually cost the company money.

The subjects for this study were a group of 60 men and women divided among two age groups and three levels of education. The subjects watched the (partially or entirely) English commercials and were asked to rate them in terms of a number of characteristics, to transcribe the English segments, to indicate whether they thought they understood them, and finally to translate the English segments into Dutch. The results showed that attitudes towards English in commercials were not very positive in any of the groups of subjects. Comprehension depended on age and level of education, but the main finding was that in two thirds of the cases, the meaning of the English commercial was not understood correctly, even though the subjects themselves may have indicated otherwise (Gerritsen et al., 2000).

It is important to keep in mind that these results apply to spoken commercials, and attitudes and comprehension may be different for other forms of communication. Other studies show similar patterns of low comprehension for English in written text (e.g. Gerritsen, 1996 and Gerritsen et al., 2010) but they also suggest that Dutch speakers have fewer problems comprehending English in written text than in a spoken format. This explains why the commercials that included written as well as spoken text in the 2000 study were understood more frequently than the others. These studies do not address anglicisms in translated texts in comparison with non-translated texts, but the characterisations of speakers' attitudes towards anglicisms do provide an indication as to the situations and text types in which anglicisms are likely to occur; English, and particularly American English, is used to give products a cool, international image (Gerritsen et al., 2000, p. 20), even though readers and viewers may not interpret it in this way.

2.7 Anglicisms in translation and other related studies

Loan words in Portuguese and English translated and original fiction

In a 2005 study, Frankenberg-Garcia investigated the use of loan words in English and Portuguese translated and original fiction. The aim of the study was to find out whether translations contain more loan words than non-translations, whether translation effaces the superimposition of languages in the source text, and whether the status of the source text's language and culture affects the use of loan words in translation (Frankenberg-Garcia 2005, p. 2).

The texts used for this study came from COMPARA, a parallel, bidirectional corpus of samples from English and Portuguese fiction. Fiction was a suitable text type because there are enough texts of this kind for each component of the corpus, and the corpus contained only published works because the process of selection and revision reduces the chance of mistakes. The samples were balanced so that they contained extracts from the beginning, middle, and ending of each book (Frankenberg-Garcia, 2009b, p. 3). All texts were less than 30 years old, although the setting was not always contemporary, and each sub-corpus contained works by several different authors and translators, although the Portuguese component of the corpus was more varied in this regard.

The identification of loan words was facilitated by the fact that COMPARA allows users to automatically retrieve foreign words from each text. However, this method only reveals loan words that have been highlighted (e.g. in italics) by the original author or translator. This means that certain words are counted as foreign in some texts but not in others. Since different speakers have different notions of what constitutes a foreign word, Frankenberg-Garcia's study is influenced by opinions of the creators of the text, and the results reflect those creators' perceptions of their own loan word use. In terms of numbers, multi-word expressions and quotations were counted as single loans, but multiple loans part of sequential lists were counted individually. The loan words were sorted by language of origin (which may be different from the language it was borrowed from).

A comparison of the average number of loan words per 10,000 words showed that translated Portuguese texts contained more loans than original Portuguese texts (over sixteen times more). In English, however, original texts contained more loans (over four times more than the translated texts). Frankenberg-Garcia also examined the presence of loan words in the translations compared to their respective source texts and found that, on average, the translation process tripled the absolute number of loans for both English and Portuguese (although this was not true for all individual texts).

For each of the two sub-corpora, Frankenberg-Garcia provides a list of the languages that the authors and translators borrow from and the total number of loans from each language. One of the conclusions is that English borrows from a wider variety of languages than Portuguese. Frankenberg-Garcia also notes that

translators into both languages "frenchified" the texts by increasing the number of borrowings from that language, which had opposite effects for Portuguese and English: in Portuguese, the French loans distanced the translations from Portuguese original texts (which contained fewer French words), while the introduction of more French loans made English translations more similar to non-translated texts.

The differences in the total numbers of loan words (i.e. from all languages) between translated and non-translated texts are so large that readers may notice a difference; perhaps the large number of loans gives Portuguese translations a more "foreign" feel than Portuguese original texts, while English readers are actually exposed to more loan words while reading original texts. The article suggests a number of possible explanations for the translators' increased use of borrowings, particularly anglicisms, including an intention on preserving the source language, an inability to find equivalents, or a lack of reticence due to English's status as a well-known, dominant language (Frankenberg-Garcia, 2005).

The findings discussed in this article that are most relevant to this thesis are the differences in loans from English-Portuguese language pair itself, particularly the differences between translational and non-translational texts in each language. At first glance, the data seem to suggest that translated texts are significantly richer in loan words from their source language than comparable original text: the analysis of the corpus of Portuguese original texts revealed 22 English loan words across 2 texts, while the translated Portuguese texts was found to contain 375 loans across 13 texts. English translations also showed more Portuguese loan words than English original texts, although the difference was much smaller (35 Portuguese loans in 7 of the translated texts and 14 loans in 1 of the original texts). These numbers suggest first of all that Portuguese translations are more permeable to loans from the source text than English translations and moreover that translations in general are more likely to contain loan words than non-translational texts. However, the method used to identify the anglicisms in these corpora may be part of the reason behind these results: the translated texts do not necessarily contain more loans, only more words that were *marked* as loans. The corpus of original texts could contain a significant number of loan words that went unidentified because the author did not feel the need to highlight them through their use of punctuation. It is not unthinkable that due to their experience with the relationship between source and target languages, translators are more aware of the presence of loan words and consequently more likely to foreground them as such. Taking into account the limitations posed by the method of anglicism identification, the study by Frankenberg-Garcia provides an indication of the typical characteristics of loan words in translation, but the findings regarding English loan words in the translational and non-translational corpora will not necessarily apply to other studies on anglicisms in translation.

English influence on Italian translations of articles related to business and economics

Musacchio (2005) examined the influence of English on Italian in the field of business. The aim of her study was to look beyond lexical borrowing and to determine the extent to which language contact in translating affects the target text in terms of transfer of patterns, e.g. syntactic constructs, cohesion, and reproduction of source text repetition. Due to its productive nature, structural influence can be hard to trace. Upon close inspection, Musacchio notes, syntactic loan constructions often turn out to be pre-existing native constructs that have experienced a boost as a result of language contact. Despite the uncertain origin of some of the loan constructions, Musacchio's study gives an insight into the types of influence English has beyond the lexical level.

The corpus selected for this study consisted of original English texts, their Italian translations (i.e. the parallel component), and original Italian texts (i.e. the comparable component). The corpus was intended to represent a specific language so that it contained easily identifiable terminology and phraseology while also consisting of texts that were directed at as wide a readership as possible within their field. In terms of the text type and source, Musacchio ruled out journal articles and university textbooks because their intended readership is too limited, and decided on newspaper articles instead.

Musacchio's method was to first analyse text and sentence length in order to identify English influence at macro level and then to study the corpus at micro level to determine English influence on lexis, syntax, and Baker's (1993) six translation universals. First, she analysed the corpora using WordSmith Tools to extract loan words and to determine sentence length and total length in relation to the number of tokens. Second, she compared the borrowings from English in the parallel corpus and contrasted them with the comparable texts in order to detect the influence on word formation through compounding and derivation. Third, she investigated the translation universals identified by Baker by comparing the source and target texts and contrasting them with Italian original texts. Concordancing software allowed for repetition and cohesion to be analysed automatically to a certain extent. Musacchio also compared the results with data from another English-Italian corpus of economics.

The analysis at macro level revealed that Italian translations tend to be longer than their source texts. At the sentence level, however, the average sentence length of some translations was lower than that of original texts. This difference may be caused in part by the insertion of subheadings and short sentences for marked contrast. Musacchio notes that Italian generally prefers longer, more complex sentences, but that there has been a trend towards shorter sentences, possibly due to British and American influence. Perhaps these translations reflect that trend.

In terms of lexis, the percentage of borrowings is lower in the parallel component of the corpus than in the comparable component, i.e. the translational Italian texts contain fewer borrowings than the original Italian texts. A comparison

with an Italian reference corpus called Surrey-Trieste shows that the latter contains an even lower percentage of loans—most likely due to the anti-borrowing policy the texts in this corpus are subject to. The most common types of borrowing are single-word and compound terms. The hybrid forms tend to follow the Italian word formation model.

Musacchio discusses the corpus in terms of five of Baker's translation universals (1993) (leaving out naturalisation, "which by definition excludes possibility of the influence of a foreign language"). Explicitation is sometimes sparked by foreign words in the parallel corpus where the translator feels the need to explain the term when it is translated literally. Explicitation also occurs in the form of added cohesive devices. Simplification occurs in the form of omission, for instance due to the different nature of the two languages at play that mean some source language elements would be considered redundant in the target language. Normalisation mostly applies to word order and creative language use. Repetition is often avoided in Italian (unless it gives rise to ambiguity) and replaced by synonyms, hypernyms, metonyms, ellipsis, paraphrase, or other forms of reiteration. Finally, certain features that are more common in English than in Italian may be copied into the translation, e.g. the use of a demonstrative pronoun without the addition of a noun for textual linkage which is more typical of English than of Italian. All these features of translational Italian show that English influence on Italian is not restricted to lexical borrowing but also results in the transfer of patterns (Musacchio, 2005).

Anglicisms in English and Italian business discourse

Sara Laviosa's 2007 article on studying anglicisms with parallel and comparable corpora also examines English influence on Italian. Where Musacchio investigated the transfer of patterns, Laviosa focuses on the lexical level, analysing the use of anglicisms in cross-linguistic and inter-linguistic business communication in English and Italian. The texts analysed for this study were found in a special purpose corpus consisting of two components: one English-Italian comparable corpus called ComIC&ComEC, which represented cross-linguistic communication, and one unidirectional English-Italian parallel corpus called BusiPC, which represented cross-linguistic communication.

For the identification of loans, Laviosa refers to Görlach's definition mentioned in section 2.3, which characterises anglicisms as recognizably English in their form, but accepted as items in the vocabulary of the receptor language (Görlach, 2003, p. 1). This definition excludes instances of ad hoc, transient loans and focuses on words that have at least been accepted by a group of the language's speakers. Laviosa retrieved all anglicisms from the corpus by identifying them in word frequency lists. She then produced sets of English-Italian comparable concordances for all items to find their characteristics in terms of collocation, colligation, semantic preference, and semantic prosody. The aim was to analyse the extent to which anglicisms are "functionally complete units of meaning", i.e. whether they form units

that can be compared across languages in terms of denotation, connotation, and pragmatics.

To answer this question, Laviosa specifically discusses the word *business*, which is a well-established anglicism in Italian and the most frequent English word in the ComIC corpus. Laviosa's analysis of this word in ComIC&ComEC in terms of collocation, colligation, semantic preference, and semantic prosody unveiled four comparable units of meaning for this particular anglicism. Further investigation of the Italian component of the corpus yielded a number of native equivalents for three of these senses, several of which show a tendency towards paraphrasing as a form of explicitation. Additionally, the concordances showed that the word only tends to be translated with *business* when referring to a particular economic activity. It does not replace native words but it "wedges itself into an existing semantic field" and serves as a differentiator, taking over a range of denotations that are also expressed by native equivalents (Laviosa, 2007).

2.8 Conclusion & hypotheses

Borrowing words from the source text seems to be accepted as a translation procedure, as long as it is applied judiciously with consideration of the text's readership and stylistic function. Still, translators' awareness of their role as a mediator may lead them to choose a native translation where writers of original texts would opt for anglicisms. The three general tendencies that translators seem to have to explicate, simplify, and conventionalise all have the potential of affecting their decision to borrow, as all three of them seem to favour interpretative, target-language-oriented translations over foreign words (i.e. items transferred from the source language). Based on these translation universals, it seems that the answer to the research questions should be that the translated texts contain fewer anglicisms because they are replaced with clearer and/or more explicit native terms, and that translators' tendency to conventionalise will limit the range of anglicisms they use. When considering Toury's law of interference, on the other hand, it seems reasonable that Dutch would be receptive to interference from English as a dominant, prestigious language. Nevertheless, the effect of prestige should also apply to non-translational text, and perhaps even more so, since their authors may not have the same reservations towards borrowing that translators do, so the original texts should contain at least as many anglicisms as the translated texts.

The data from Frankenberg-Garcia's study on the loan words in Portuguese and English translations and original texts showed that the translators borrowed more source language words than the original texts. In Musacchio's study of English loans in Italian, however, the translations contained fewer borrowings than the original texts. Judging from these results, no clear trend on translator's usage of loan words seems to exist so far. Moreover, these studies both cover very different types of texts—fiction and business discourse—and the food and recipe texts analysed in the chapters below are of a different type still. The different language pairs may also

influence the process of borrowing in translation; Dutch is more closely related to English than the Romance languages in the studies by Frankenberg-Garcia and Musacchio, and in combination with the dominant position of English over Dutch, this may increase the chance of borrowing. On the one hand, the differences in languages and text types may mean that the studies described are too different to be compared in terms of results. On the other hand, the methods for the analysis of borrowing in translations and original texts using corpora can be applied universally across languages (as long as there are written texts that can be analysed digitally) and comparing different languages and text types allows users of corpora to test the translation universals introduced in section 2.4 in a variety of situations.

Chapter 3: Methodology

3.1 Introduction

The method of investigating anglicisms for this thesis involved the selection and analysis of a comparable corpus made up of samples from cookbooks. This chapter addresses the practical aspects of text selection, corpus compilation and processing, and analysis of the anglicisms. The results of this analysis will be presented in chapter 4.

3.2 Selection of the corpus

Text type and genre

The decision on cookbooks as the source for this corpus was based on a number of factors. First, cookbooks are usually made up of two types of text: the recipes themselves, which tend to follow a conventional pattern that is quite similar across different books, and the introductions and head notes, in which the author writes freely about topics related to the food and the stories behind it. This combination of typical standard phraseology and informal, conversational writing should produce a corpus that contains conventionalised anglicisms as well as more spontaneous borrowings. Second, it is likely that at least some of the cookbooks published in a country reflect the current trends in that society (e.g. diets and food fads). Since English represents fashionableness in Dutch (Ridder, 1995, p. 48), this seems like a genre that would be receptive to anglicisms. The corpus is likely to be limited in size due to time constraints, so selecting texts that are rich in borrowings in order to obtain as much data as possible seems like an efficient choice. Third, the number of cookbooks that appear in Dutch every year is limited enough that it is possible to select a representative sample relatively easily and large enough to still form a corpus that is varied in terms of authors, translators, and cuisines.

Other sources that were taken into consideration to be part of the corpus were online magazines, newspapers, and blogs, particularly in the categories of food and lifestyle (for the reason mentioned above). Depending on the topic, these sources also have the degree of informality that would make them receptive to anglicisms, and an obvious benefit is that the texts are already digitalised and ready to be analysed by corpus software. However, the nature of these sources makes it difficult to select a comparable corpus of translated and non-translated texts: most articles on Dutch websites seem to be original Dutch texts and if they are translated or adapted from an English article, this source is not always stated. Blogs are even more problematic in this regard because the writer and translator may be the same person, which means the relationship between source and target text becomes unclear. Since publishing online is such an informal process, a corpus compiled from these texts

could provide interesting insights into Dutch speakers' "natural" tendencies in their use of English, but it was found to be impractical for the purpose of this thesis.

Compiling a corpus only from published books removes a lot of the problems associated with online texts, since publishers usually clearly state the writer, translator, and original title of their works, so selecting texts for both the translational and non-translational sub-corpora is relatively straightforward. Zanettin (2002) also argues in favour of the use of published books because they are considered to be "central to accepted standards of language production" and the standardised editing process reduces the occurrence of mistakes. Bestseller lists provided by book sellers indicate which texts can be used as representative for a particular period. In terms of anglicisms, publishers may have overt or covert policies that determine the way in which borrowing is represented in their works. This may or may not be favourable for the analysis of their texts: on the one hand, policies and editing processes may limit the number of loans that make it into the final text so that it does not reflect the authors' own writing; on the other hand, the anglicisms that do end up in the final text may be said to be representative of what is considered acceptable and "standard" in the target language.

Another alternative method would have been to use comparable corpora constructed by others. This could certainly have saved time by eliminating the compilation process, but it would have imposed a number of limitations. The main problem is related to the identification of borrowed elements, which are not necessarily labelled as such in existing corpora. Even if they are, the user is dependent on the compilers' or authors' definition of anglicisms; in the COMPARA corpus described in section 2.7, words were only counted as anglicisms if they were highlighted as such in the original texts. This criterion seems a little arbitrary, and it could lead to the exclusion of a significant number of relevant anglicisms. A way of resolving this issue could be to use texts from existing corpora, but to ignore any existing tags and to identify the loans manually. However, many corpora do not seem to offer access to the full texts and only allow users to perform concordance searches of the texts using an online interface. Finally, as Zanettin (2002) pointed out, comparable corpora are not necessarily representative of the original texts published in one of the languages they contain, because they consist only of texts that have been translated into another language.

Selection and digitalising of texts

All in all, compiling a special purpose corpus for this thesis seems like the most effective approach here. This method allows the user to customise a corpus to the specific requirements of their research questions. As with other types of corpora, the quality and size of these ad hoc corpora are limited by time and resources, and if a corpus is only used once, it is especially important to make the process as efficient as possible. This may mean the corpus is limited in size, but as Bowker and Pearson (2002) point out, sometimes "you can get more useful information from a corpus that

is small but well designed than from one that is larger but not designed to meet your needs". The time restrictions mean that some form of compromise seems unavoidable, but if these limitations are taken into account during the analysis of the results, the data from the corpus can still be used effectively.

Bowker and Pearson suggest starting the selection process by describing the ideal version of the imagined corpus in terms of size, number of texts, medium, subject, text type, authorship, language, and publication date (Bowker & Pearson, 2002, p. 69). This technique is intended to make the compilation process more efficient by removing all irrelevant texts from the compiler's consideration. The main demand on the corpus "wish list" for this thesis was that it needed to contain both texts written originally in Dutch (the NL-OR sub-corpus, for short) and texts that were originally English and translated into Dutch (the NL-TR sub-corpus). The other features on the list that the texts should be written digital texts on the subject of food, published in the past ten years, in the form of full texts or relatively large samples (i.e. 20-25 pages), written by a variety of authors and translators (starting at 20 with the option of expanding later on). Most of these demands had to be compromised to a certain extent during the compilation process, mostly due to limited availability of digital texts, but all of the features are present in the final corpus to some degree: the final corpus consisted of 54 cookbook excerpts, half of which was originally written in Dutch and half of which was translated from English, all published within the past ten years and written by different authors (though a few translators occurred twice).

The books included in the corpus were selected using two bestseller lists available online: first, the archives of the food and drink section of CPNB's weekly *Bestseller 60*—which lists bestselling books based on information obtained from over 900 Dutch book shops—from 2012, 2013, and 2014 ("De Bestseller 60", 2014), and second, online book seller Bol.com's section on bestselling cookbooks—which is updated daily—of 29 April, 2014 ("bol.com | Bestverkochte kookboeken", 2014). Once a number of "candidate texts" for the corpus had been accumulated in the form of a list of recent popular cookbooks, the next step was to select texts to sample. This decision was mainly based on availability: most publishers offer some type of preview of their books online, but not all of these were equally suitable for corpus analysis. PDF files or other types of selectable text were given preference because the text could be copied and pasted into text files and analysed using corpus analysis software without needing much further processing.

Some publishers only offered previews in the form of images, while others offered no online previews at all. These texts required a number of extra steps in order to be made analysable, but it still seemed important to include these texts in the interest of creating a representative corpus—otherwise the corpus would only reflect the publishers that chose to publish their texts in a digital format, perhaps to the exclusion of more traditional publishers. The texts that were only available as images were converted into text using an optical character recognition tool (TopOCR) and other texts were digitalised by first scanning the pages from printed books and then converting them using the text recognition software. The use of full

books offered the benefit of being able to select a more representative sample both in terms of size and composition (i.e. a fixed number of pages from the introduction as well as other sections of the book), but the process is quite time-consuming. Despite the attempt to include texts from different publishers and formats, the final corpus is still largely determined by availability: some books offered no online previews, some texts were unsuitable for conversion using OCR tools due to irregular backgrounds or small print, and the use of printed books was limited by the availability of titles at the library.

To ensure that each text formed a representative sample of the book it was extracted from, samples were taken both from the introduction and from different sections of the recipe component of each book. In cases where the online preview restricted the number of available pages, samples were taken from a more limited number of sections, but all texts are made up of a combination of both general and instructional texts. The details of the texts that comprise each corpus (including the titles, authors, and number of words per excerpt) can be found in Appendices 1A and 1B.

In addition to the comparable corpus of Dutch original and translated texts, a smaller corpus was compiled from the texts that formed the source texts of the NL-TR corpus. The reason this corpus (EN-OR for short) is smaller than the other two is that the excerpts that were available for the books of the source and target texts only overlapped to a certain extent, so not all of the text from the NL-TR corpus could be linked to its source text. Even in its limited form, however, the EN-OR corpus can be used to analyse the translators' use of anglicisms in more detail and to help explain why they choose to borrow some words and not others.

Size and representativeness

Many corpus-based studies rely on the size and representativeness of their corpus for their results to be relevant, but other than "more is better" (Baker, 2007, p. 52) no clear consensus on the topic exists (Corpas Pastor & Seghiri, 2009). As a result, corpus size is too often determined by availability of texts rather clear criteria. Corpas Pastor and Seghiri introduce a method that determines the representativeness of a corpus by monitoring the type/token ratio as the corpus size increases. This ratio is likely to be high at the beginning of the compilation process when the corpus contains few words, so that a relatively high number of new types are introduced with each additional text, but once the total number of words increases and the chance of new words being introduced goes down, the ratio should drop rapidly. The authors argue that a corpus can be considered to be representative when the addition of new words has little to no effect on the overall type/token ratio.

Corpas Pastor & Seghiri demonstrate their method with graphs made using the ReCor software (figure 3.1 below). In these graphs, the horizontal axis represents the total size of the corpus (either in documents, for graph A, or in tokens, for graph B), while the vertical axis shows the type/token ratio. The documents are entered

both alphabetically and at random (represented by the two separate lines) in order to ensure that the order of introduction does not influence the results. When both lines stabilise as they approach zero, the introduction of new corpus no longer significantly affects the type/token ratio and the corpus can be considered representative of the selected genre (Corpas Pastor & Seghiri, 2007).
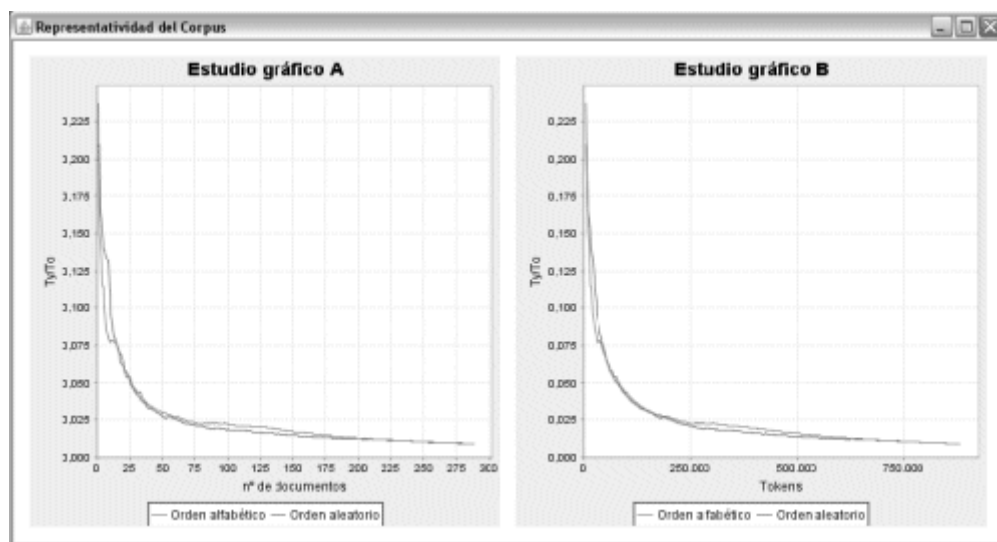
Figure 3.1: Graphs demonstrating the representativeness of a corpus (Corpas Pastor & Seghiri 2007)

The software used to generate the above graph currently seems to be unavailable, but the same principle can be applied by manually splitting the corpus into sections of equal size, adding these files to the Wordlister function of a corpus processing tool and keeping track of the type/token ratio in between additions. This method yielded the graphs for the NL-OR and NL-TR corpora shown below in figures 3.2 and 3.3.
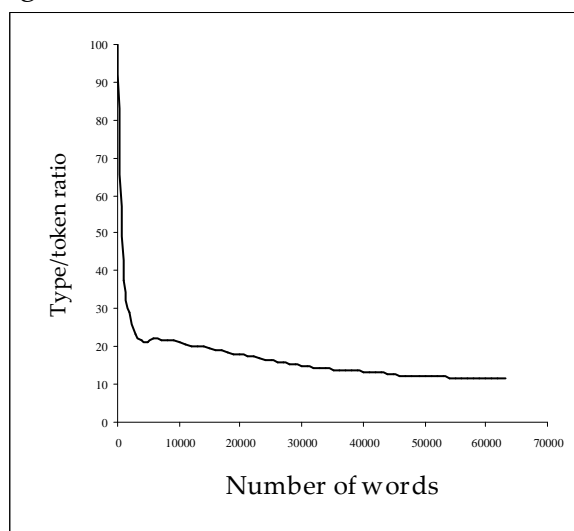
Figure 3.2: The type/token ratio of the NL-OR corpus as total size increases
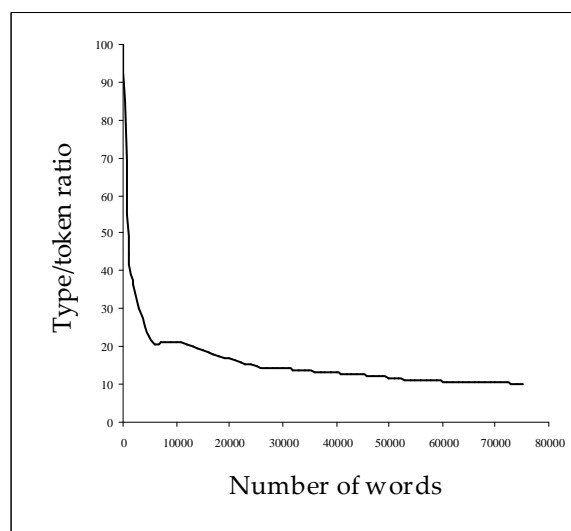
Figure 3.3: The type/token ratio of the NL-TR corpus as total size increases

The graphs illustrate how the type/token ratio for each corpus goes down rapidly with the addition of the first few texts and begins to stabilise near the end of the graph. Both corpora would still benefit from the addition of more texts to add a wider variety of authors and data (which should make the graph stabilise even more visibly), but this method suggests that they are at least usable in terms of size.

In addition to the number of words and documents, the representativeness of a corpus is also determined by the characteristics of its components. Halverson (1998) suggests that a representative corpus may be centred around professional translations and contain additional sub-corpora with related texts (e.g. translations by beginning translators, second language speakers, etc.) with varying degrees of significance and relevance which are "all being regarded as legitimate objects of study" (Laviosa, 2002). The corpora assembled for this thesis are not necessarily as varied as Halverson advises in terms of the level of translators' experience (the fact that these are popular books published by well-known publishers indicates at least a certain degree of professionalism), but the texts in the corpora do vary in terms of the different cuisines and diets they cover, and this difference could be used to analyse the relationship between translation and lexical borrowing in terms of the various subgenres as well.

Ideally, all components of a corpus should be the same size (e.g. 5000 words per text), so that the data extracted from the corpus can be said to be representative of all of the texts. However, the different sizes of the excerpts available for the cookbook corpus resulted in a collection of texts that varied widely in size. There seems to be no ideal solution to this problem. The possible ways to balance the corpus are to either reduce all texts to match the size of the smallest one (i.e. cutting down all texts, including the ones made up of 5000 words or more, to a mere 432 words) or to simply exclude all smaller texts (which would result in the exclusion of almost all texts obtained through online previews and greatly reduce the variety authors and translators). Clearly, both of these methods would result in a significant loss of data. For this reason, the corpora for this thesis were composed of texts of varying sizes. The consequence of this decision is that the resulting corpus is unbalanced. This does not necessarily pose a problem for the analysis of the results, as long as the imbalance is taken into account. In order to ensure that frequency data was not affected by the overrepresentation of individual authors and translators, the average number of loans per 1000 words was calculated for each text before using these numbers to identify tendencies of the authors and translators in general.

3.3 Anglicism identification and frequency

Defining and identifying anglicisms

Because the corpus was composed ad hoc and not tagged in any way, the most effective way of extracting a list of anglicisms was to go through the texts and identify the loans manually. The use of text recognition software already made it

necessary to check all texts for correctness, so identifying anglicisms at the same time did not require much added effort. For a larger corpus, however, this method may be too time-consuming and impractical. Other possible methods are to identify the anglicisms frequency lists (though seeing the terms outside of context may make it more difficult to recognise them as borrowings) or to start with a small, representative section of the corpus, to identify its anglicisms, and to use the resulting list to analyse the use of these terms in the rest of the corpus.

As exemplified in section 2.3, definitions of what constitutes an anglicism may be very broad (i.e. any feature that is in some way influenced by English) or quite restricted (i.e. only words that are recognisably English in form and generally accepted by recipient language speakers). The analysis below is limited to lexical items that have entered the Dutch language through English. This definition includes ad hoc loans that have not necessarily been integrated or accepted by the majority of speakers as well as loan translations and other conventional borrowings that speakers may not directly recognise as English, but it excludes structural borrowing. The reason for including ad hoc loans in addition to generally accepted anglicisms is that these are likely to be the most salient; because, by nature, unconventionalised loans stand out more than integrated terms, they are likely to leave more of an impression on the reader, and excluding them from this analysis would not provide an accurate representation of anglicisms in translations and original texts. The reason for the exclusion of structural borrowing is that lexical items can be identified as borrowings quite easily—in case of doubt, the online *Etymologiebank* (Van der Sijs 2010) was used for reference—but the origin of grammatical structures is more difficult to trace (as mentioned in the description of Musacchio's 2005 study in 2.7). Structural borrowing was investigated to a certain extent by using a list of commonly occurring loan translations (as described in 3.5 and 4.4), but lexical items had the main focus.

The method of manual selection means that it is possible for anglicisms in individual texts to have gone unnoticed, but the subsequent concordance searches of the entire corpus ensured that at least all occurrences of the most frequently occurring terms were counted. The software used to analyse the corpora for this thesis was Corsis (formerly called Tenka Text), which includes both a wordlister and a concordancing tool. As mentioned above, an effective method of expanding the corpus would be to use the list of loan words from the first corpus and to apply it to an expanded corpus. Assuming that the texts are similar enough that the most frequently occurring loans will be more or less the same, the data from the larger corpus could be used to verify and strengthen the results obtained from the first corpus.

Frequency

Once a list of anglicisms had been compiled manually, concordancing software was used to determine exact number of times each term occurred in each sub-corpus.

During this process, different Dutch hybrids formed using the same English borrowing were classified as multiple occurrences of the same anglicism (i.e. multiple tokens of the same type). For example, *gezinslunch*, *lievelingslunches*, and *meeneemlunch* were all counted as instances of the anglicism *lunch*. If an anglicism consisted of two English loans that also occurred separately, however, the compound was counted twice: *custardpuddinkje* was counted as a variation of *custard* as well as *pudding*.

In order to provide as accurate an indication as possible of the frequency of anglicisms in each sub-corpus, the average number of anglicisms per 1000 words was calculated for each text in the corpus. These averages determined the total frequency of anglicisms for the sub-corpus they form. Because some excerpts are over ten times larger than others (e.g. Ottolenghi's *Plenty* and Paltrow's *(H)eerlijk* in the NL-TR corpus), the danger that one text's idiosyncrasies may overpower the other texts should be taken into account. Calculating the averages separately before determining the total frequency ensured that no one author's tendencies were overrepresented.

3.4 Classification of the anglicisms

Type

As described in 2.3, Haugen (1950) suggests that all loan words can be classified in terms of two factors—the elements that are imported from the donor language and the elements that are substituted from the recipient language. The combination of these factors results in three main categories under which the anglicisms from the cookbook corpus can be filed: loan words (e.g. *bagel*), loan blends (e.g. *muffinbakjes*), and loan shifts (e.g. *supermarkt*). The label *loan word* may be somewhat misleading because the category also includes anglicisms that consist of more than one word, but for the sake of consistency the analysis below will stick to Haugen's terminology. The decision to use Haugen's typology rather than, for instance, Gottlieb's, is based on its simplicity; the three categories all apply to lexical borrowing and even for a relatively small corpus, they provide a broad indication of the composition of its anglicisms. In addition to these three types, there was a small group of anglicisms that appear to have been coined under the influence of English but do not exist in the donor language itself; these words were classified as loan creations.

Grammatical category

As described under 2.3, Muysken suggests that there are certain parts of speech that speakers are more likely to borrow than others. In order to see whether there are any differences between translators and writers in this regard, all occurrences of the anglicisms found in the corpora were classified based on grammatical category (nouns, verbs, adjectives, adverbials, and full phrases or sentences). The resulting percentages were used to determine whether these hierarchies of borrowability

apply to the translated texts in general and particularly in comparison with non-translational text.

Function

The functions of anglicism use introduced by Onysko (2004) and Galinsky (1967) as described in section 2.3 included semantic, stylistic, euphemistic, emotive, or social motivations as well as the purposes of brevity, precision, satire, vividness, and painting an American setting. These typologies provide a useful indication of the different situations in which speakers are likely to use anglicisms, and many of these functions are likely to apply to the anglicisms analysed in this thesis as well. However, when looking at the purpose of anglicisms in practice, it seems that writers often have more than one reason to borrow and Galinsky and Onysko's motivations may overlap. For instance, an anglicism may be shorter as well as more specific in meaning than its native counterpart, and an author may choose to use this word for its associated prestige as well as socially to signal their group identity. Moreover, it is difficult to determine the speaker's true reasons just by analysing the text, and even if it were possible to ask them about their motivations, they may not be able to give an objective answer themselves.

The main frame of reference here is the text itself, and what seems to be the most sensible method is to identify native equivalents for each anglicism and then to analyse how the instances of that loan compare to the native terms in terms of Onysko and Galinksy's suggested motivations. Prestige appears to be a factor that is always present in varying degrees, so classifying anglicisms as either motivated by prestige or not motivated by prestige would be ineffective. However, there are cases in which none of the other motivations seem to apply and prestige seems to be the primary factor influencing the borrowing. During the analysis, the range of functions fulfilled by the anglicisms was essentially condensed into two main motivations: semantic (i.e. for precision, e.g. so-called "cultural borrowings" which have no native equivalents or words that have different connotations from the existing native words) and stylistic (e.g. anglicisms that are used for brevity, alliteration, variation of expression, or to add a foreign atmosphere). Even these two categories are bound to overlap; it may be argued that all borrowings are motivated by semantics to a certain degree, because even its status as an anglicism is likely to add a certain connotation to a term. In cases where the borrowing clearly served a stylistic purpose (e.g. because it is much shorter than native equivalents) it was included in the latter category.

3.5 Conventionality of the anglicisms

In an attempt to compare the sub-corpora in terms of the conventionality of their borrowings, the anglicisms in each sub-corpus were first classified in terms of their inclusion in (or exclusion from) Manfred Görlach's Dictionary of European

Anglicisms. As mentioned in 2.3, this dictionary only includes anglicisms that have been accepted by a significant group of a language's speakers. To approach the concept of conventionality from a different angle, the anglicisms were also entered into a reference corpus of contemporary Dutch. Assuming that translations tend to be more conventional than non-translated texts, as suggested by Kruger (2002) and other researchers mentioned in section 2.4, the list of anglicisms found in the NL-TR corpus should consist of a higher percentage of loans that are included in these reference works than the anglicisms in the NL-OR corpus.

The two reference works described above provide a very general indication of the anglicisms' degree of conventionality. To examine the conventionality of more specific types of borrowing, the anglicisms were analysed in terms of their adherence to conventional Dutch spelling rules. The expectation was that the authors of original texts would be less resistant to unconventional spellings than the translated texts.

3.6 The comparable corpus

As described in 3.2, a small parallel corpus of source texts (EN-OR) was compiled in order to analyse the translators' use of anglicisms in more detail. This corpus consisted of seventeen excerpts (the other ten books had to be excluded due to unavailability). The source text corpus was not linked to the comparable corpus using parallel corpus software; the different composition of the source and target texts in addition to time restrictions made this impractical, and since only the borrowed words were relevant, it seemed unnecessary to align the entire corpus. Despite the limitations of the corpus, it still yielded a useful list of anglicisms and their source text terms as they occurred within the original context.

3.7 Conclusion

In short, the main steps taken to answer the research questions for this thesis were (1) to decide on a text type and to make a selection of texts; (2) to extract samples from these texts and edit them so they could be analysed using corpus processing software; (3) to identify all anglicisms in the corpus and to determine the frequency for each individual text as well as the averages for both of the sub-corpora; (4) to classify all anglicisms based on type, grammatical category, and function; (5) to compare the anglicisms in the corpus to reference works and official spelling rules in order to establish the degree of conventionality; and (6) to link the anglicisms in the NL-TR corpus to the original terms in the EN-OR corpus and to analyse the situations in which translators borrow in more detail.

Chapter 4: Results

4.1. Introduction

This chapter summarises the data that were extracted from the cookbook corpus described in 3.2 using the methods described from 3.3 onwards. The main focus in this chapter is on the different proportions in the NL-OR and NL-TR corpora in terms of the frequency, types, functions, forms, and degrees of conventionality of the anglicisms (sections 4.2-4.4). Section 4.5 describes the findings derived from the analysis of the NL-TR corpus in comparison to the EN-OR corpus of source texts. Where relevant, the results are summarised in tables and illustrated with examples of citations from the corpus. The further interpretation and relevance of the data will be discussed in Chapter 5.

4.2 Frequency

The table below shows the average number of anglicisms per 1000 words for each separate text in the NL-OR corpus in terms of the total number of occurrences (tokens) as well as the different types of anglicisms.

Table 4.1: The average number of anglicisms per 1000 words for each text in the NL-OR corpus

Title	Anglicisms per 1000 words (tokens)	Anglicisms per 1000 words (types)
70 Groene Smoothies	48.5	8.3
Boekoe Bangsa	11.8	10.6
Comfort Food	7.7	6.8
De Dunne Vegetariër	5.9	4.2
DedikkevanDam	5.2	4.7
Down to Earth	14.7	8.7
Ellemieke's Familie Kookboek	27.5	9.5
Eten uit de Natuur	9.5	5.6
Green Delicious	2.6	2.6
Grenzeloos Koken	6.2	4.5
Het Beste Dieet van de Wereld	9.4	9.4
Home Made Zomer	23.7	15.1
Impress Your Friends	13.0	5.8
Koken met Kennis	19.5	10.7
Lekker Lang Jong	10.3	3.3
Oergondisch Genieten	8.9	4.6
Puur Genieten 2 En Toch Gezond en Slank	14.3	2.2
Rudolph's Cupcakes Compleet	67.0	16.9
Rutger Bakt	18.6	7.4
Smart Cooking Compleet	8.4	4.6

Superfood Recepten	41.3	7.6
The Culy Way of Life	23.8	18.6
Werken met Vis	1.7	1.7
Winterslank	9.9	2.2
Yoga Kookboek	12.6	6.3
Yvestown in de Keuken	19.2	14.0
Koken met Karin Zónder Pakjes en Zakjes 2	12.9	4.0
Total average	**16.8**	**7.4**

The table below displays the same information for the NL-TR corpus.

Table 4.2: The average number of anglicisms per 1000 words for each text in the NL-TR corpus

Title	Anglicisms per 1000 words (tokens)	Anglicisms per 1000 words (types)
(H)eerlijk	25.2	13.3
30 Minuten Vegetarisch	19.5	9.8
365 Sappen & Smoothies	18.4	3.5
500 Stoof- & Ovenschotels	6.1	3.4
500 Sushi	3.1	3.1
500 Taarten en Cakes	20.9	7.4
Baking Made Easy	16.2	5.6
Bakken met de Cake Boss	15.0	10.0
Chez Rachel	10.9	5.2
De Echte Chinese Keuken Thuis	3.5	1.7
De Free Range Cook	19.4	6.0
De Kunst van het Bakken	13.2	8.3
De Smaken van Spanje	1.9	1.9
De Souk	4.4	2.4
Delicious. Lekker Koken!	16.8	8.9
De Basics	13.1	4,8
Heston Blumenthal Thuis	12.7	7.8
Het Grote Granenboek	8.5	4.4
Italiaanse Hapjes	0.0	0.0
Jamie in 15 Minuten	17.1	4.6
Koken met Quinoa	5.9	3.0
Matt Preston's 100 Beste Recepten	18.0	10.8
Nigellissima	7.1	4.4
Plenty	6.7	2.8
Raw Food als Levenskunst	6.9	6.9
Roken, Drogen en Pekelen	16.8	5.9
Veg!	4.9	3.4
Total average	**11.6**	**5.5**

The total average occurrences of anglicisms per 1000 words for the two sub-corpora are not widely apart (16.6 for NL-OR and 11.6 for NL-TR), but the difference is clear enough to suggest that the non-translational texts are generally richer in anglicisms. The average number of types suggests the same: the authors of the original texts draw from a wider variety of anglicisms, whereas the range of different anglicisms used by translators seems more restrictive. The NL-OR corpus contains more texts that are exceptionally rich in anglicisms (6 out of 27 original texts and 2 out of 27 translated texts contain over 20 anglicisms per 1000 words) and the NL-TR corpus consists of more texts that contain relatively few anglicisms (6 of the translated texts and 2 of the original texts contain fewer than 5 anglicisms per 1000 words). The NL-TR corpus even contains a text with no anglicisms at all, though this may be explained by the fact that the excerpt is rather small.

The finding that the translators seem more reluctant to use anglicisms than the Dutch cookbook writers is consistent with the translation universals discussed in section 2.4 that suggest that translators are more inclined towards explicitation and conventionalisation of their source texts. It is also in line with Musacchio's finding that Italian translational texts in the field of business and economics contain fewer anglicisms than comparable non-translational texts. A comparison of the percentages suggest that there is something about the Dutch cookbook corpus that makes it more susceptible to borrowing than the texts selected for Musacchio's corpus: the sub-corpus with the most anglicisms in her study contains about 8.5 lexical borrowings per 1000 words, which is lower than most of the texts in the cookbook corpus. This difference may be caused by the different subject matter (e.g. the relative informality of food writing), by the different language pairs at play (the fact that English is more closely related to Dutch than to Italian may facilitate borrowing), or by a combination of these and other factors. The results in the above table seem to go against the results from Frankenberg-Garcia's study that showed that translations are more likely to contain loan words from the source text than comparable original texts in the target language. However, as described in 2.7, the universality of this study's findings may be limited by the method that was used to identify the anglicisms.

4.3 Classification

Type

In both sub-corpora, pure loan words (i.e. borrowings that display importation from English and no substitution from Dutch) account for over half of all occurrences of anglicisms. Loan blends seem to be more common in the NL-TR corpus than in the NL-OR corpus (35.9% vs. 29.4%), but the method of counting these proved to be somewhat problematic due to the occurrence of some ambiguous nouns that could be classified either as loan words or loan blends. Examples of this are *avocadotoast* and *tortillawraps*, in which the second element is clearly an anglicism (Van der Sijs, 2010), but the first element could be the result of either the same instance of

borrowing from English or blending with Dutch. The orthography and common pronunciation of the terms may suggest the latter, but these may simply be a result of integration. As a result, the exact count of loan blends depends on interpretation. Loan shifts constitute about 6% of the anglicism occurrences in each sub-corpus while only being made up of four different types, three of which are very common (*biefstuk*, *diepvries*, and *supermarket*). Finally, 0.4% of the anglicisms in the NL-OR corpus are names (e.g. *24Kitchen*, *Yvestown*) in the form of loan creations, i.e. words that are English in form but have no source word in the donor language. It is unsurprising that this form does not occur in the NL-TR corpus, as most anglicisms and names are transferred from the source language, which automatically means they are not new creations.

Grammatical category

The hierarchies introduced by Muysken (2010) suggest that certain words are more susceptible to borrowing than others. The corpus used for this thesis is too small to demonstrate the effects of Muysken's hierarchies that relate to specific topics (e.g. numbers and colours, neither of which occur as anglicisms in this corpus), but there is a clear distinction between the different grammatical categories. According to Muysken, nouns are borrowed most often, followed by adjectives and verbs, and this holds true for the cookbook corpus as well. The nouns clearly form the majority with about 90% of all anglicisms (tokens) in each sub-corpus. Within this category, the NL-TR corpus contains over 1.5 times as many proper nouns as the NL-OR corpus. This count excludes the names of persons but includes names and titles of books, restaurants, brands, etc., most of which do not have native Dutch equivalents. Adjectives and verbs are the other significant groups; the original texts seem to be richer in the latter while the translated texts contain more of the former. The distribution of each category is shown in table 4.3 below. The percentages for the different grammatical categories NL-OR and NL-TR sub-corpora lie very closely together, and the corpus is likely too limited to make claims on translators' and writers' tendencies in terms of smaller grammatical groups.

Table 4.3: The distribution of the grammatical categories of the anglicisms in each corpus, in occurrences (tokens) and percentages of the total number of anglicism tokens

	NL-OR corpus		NL-TR corpus	
nouns	853	91.1%	716	89.2%
proper nouns	*35*	*3.7%*	*55*	*6.8%*
adjectives	36	3.8%	57	7.1%
verbs	35	3.7%	21	2.6%
adverbs	8	0.9%	5	0.6%
phrases / sentences	2	0.2%	4	0.5%

Function

As explained in section 3.4 above, the function of an anglicism can rarely be classified under just one category, and motivations for borrowing usually appear to be a mixture of several of the factors introduced by Onysko (2004) and Galinsky (1967). An example of this is the word *mix*: this is one of the most frequently occurring anglicisms in both corpora, and there are a number of characteristics that may explain its prevalence. First, it is shorter than its native equivalents *mengsel* and *mengen*. Second, there seems to be a semantic distinction in Dutch; in most of the cases where *mix* is used as a verb or in reference to an appliance (i.e. "mix niet te lang" or "klop met de mixer"), it specifically means combining ingredients using an electrical appliance, while the native form *meng* is used for mixing by hand. The factors of prestige and attitude are likely to be of some influence here, too—if the author or translator was very resistant to English influence on Dutch, they would likely have gone out of their way to find a Dutch equivalent. For a small number of the occurrences, prestige seems to be the main reason for the loan (i.e. occurrences of *mix* where there is no electricity is involved so the semantic distinction seems absent and where the verb is used in the imperative mood so the Dutch equivalent would be *meng*, which is almost just as short).

Overall, the functions of the anglicisms seem to be distributed similarly in the NL-OR and NL-TR corpora; around 85% of the occurrences show some kind of semantic distinction (as compared to their native equivalents) that is likely to have contributed to their use, and 6-8% seem to serve a stylistic purpose (most often brevity). The most salient difference is that the NL-OR corpus seems to contain significantly more anglicisms that are used mainly for prestige (with no apparent other motivations). This result seems like a logical consequence of the different natures of translations and original texts: while writers may choose to use a prestigious language as a way of signalling their status or group identity, translators are always speaking on behalf of someone else, and as a result, their choices are more likely to be interpreted as stylistic choices (i.e. as a way to add to the foreign atmosphere by referring to the source culture).

4.4 Conventionality

Manfred Görlach's Dictionary of European Anglicisms

Out of all anglicisms in the NL-OR and NL-TR corpora, about 65% occurred in Görlach's Dictionary of European Anglicisms. The average "degree of acceptance" for the anglicisms in both corpora was "2" (i.e. words that are accepted across a range of styles and registers but are still markedly English in form) (Görlach, 2005, p. xxiv). There was a slight difference between the two sub-corpora—63% for the NL-OR corpus and 66% for the NL-TR corpus—but it does not seem justified to say that the translated texts were significantly more conventional in terms of anglicisms

based on these numbers alone. First, the difference between the percentages simply seems too small, and second, the dictionary was published almost ten years ago and the information gathered from it is not necessarily representative of the current situation in the Netherlands. For instance, the words *website*, *magazine*, and *blog* are not included in this dictionary at all, and the word *scoren* is listed as only occurring in technical or specialist vocabulary.

Corpus Hedendaags Nederlands

Because Görlach's dictionary alone may not provide an up-to-date indication of the conventionality of anglicisms in Dutch, the *Corpus Hedendaags Nederlands* was used as an alternative resource. This reference corpus consists of over 800,000 texts from newspapers, magazines, news broadcasts, and legal writing ("Corpus Hedendaags Nederlands", 2014). Out of all anglicism tokens in the NL-OR corpus, 98% was present in the reference corpus (and 90% of its types). From the NL-TR corpus, 95% (and 85% of its types) occur in the reference corpus. These numbers suggest that the NL-TR corpus is slightly richer in highly unconventional loans. For this corpus, 50% of the anglicisms that are absent from the reference corpus are names and quotes (compared to 30% in the NL-OR corpus).

 The frequency of the different anglicisms in the reference corpus varied widely, some only yielding one and others yielding tens of thousands of hits, but there was no clear distinction between the translational and non-translational texts in this regard. The NL-OR corpus seems richer in words with 100 or fewer hits, the NL-TR corpus had slightly more anglicisms with 1,000-10,000 hits, and the NL-OR corpus contained more anglicisms with more than 10,000 hits. This comparison was made more difficult by the fact that the results cannot always be analysed based on the number of hits alone; sometimes the English word occurs within a quotation of English text, and some words occur both as an anglicism and as a native Dutch word (e.g. *plannen*, *burger*, *rare*) and these forms cannot be distinguished based on orthography alone. Manually eliminating the irrelevant entries would be too time-consuming to be practical in this case. As a result, the percentages obtained from this corpus search may not be entirely reliable, and it seems useful to look beyond mere frequency and to examine the conventionality of different word classes in more detail.

Conventionality in terms of correspondence to Onze Taal's language advice

Because comparing the degrees of conventionality of the loans in terms of their occurrence in reference works did not seem to yield much detailed or reliable information, the anglicisms were also analysed based on more specific characteristics, such as word formation and spelling. The Dutch Onze Taal website contains a number of articles that provide advice on the ways in which English words should be incorporated into Dutch ("Adviezen over Engelse woorden in het

Nederlands". n.d.). Their recommendations are based on official Dutch spelling rules (i.e. *Het Groene Boekje*) as well as alternative spellings (e.g. *Het Witte Boekje*). Determining the extent to which this advice is applied to the anglicisms in the cookbook corpus should lead to a more detailed description of the degree of conventionality in each sub-corpus.

Compounding

The feature that showed the most significant difference between the two sub-corpora was compounding. Onze Taal suggests that compounds containing English borrowings should be written as one word (so without spaces, e.g. *humanresourcesmanagement*). Hyphens may be added to improve legibility and ease of comprehension (e.g. *human-resourcesmanagement*) and are required if they were already present in the English compound (e.g. *all-invakantie* or *down-to-earthbenadering*) ("Engelse woorden in Nederlandse samenstellingen", n.d.). The total percentages of compound nouns were 20.8% of tokens in the NL-OR corpus and 31% in the NL-TR corpus. For the NL-OR corpus, 3.6% of these compounds contained non-obligatory hyphenation (e.g. *smoothie-boek, courgette-gehaktburgertjes*) and 3.6% contained spaces rather than being written as one word (e.g. *pastinaak chips, cream cheesecrème*); for the NL-TR corpus, these percentages were 10.8% and 2.8%, respectively. In other words: (1) the NL-TR corpus contained more compounds in general, (2) the NL-TR corpus contained more words that were deliberately hyphenated, and (3) the NL-OR corpus contained a slightly higher percentage of compounds with superfluous spaces.

      In terms of conventionality, these results may be interpreted to mean that the OR-NL corpus adheres less to the "accepted" spelling due to its unconventional use of spaces, but the difference between the percentages is quite small. The proportions are slightly clearer—i.e. 2.6% vs. 1.6%—when compounds consisting only of potentially English elements are disregarded (e.g. *mango chutney* and *lemon curd*, which may just have been borrowed without adaptation to Dutch spelling rules), although a larger corpus would be needed to determine whether there truly is a significant difference here. The difference in the frequency of hyphenation is more pronounced; optional hyphens are more prevalent in compounds in the NL-TR corpus. This does not necessarily mean this sub-corpus is less conventional (Onze Taal considers extra hyphenation acceptable and sometimes even desirable), but it does suggest that the translators made an extra effort to make the words easier to understand: perhaps this is an example of translators' tendency towards explicitation.

Diminutives

Onze Taal also has articles on the spelling of diminutive forms derived from English loans. The advice here is that the spelling should conform to the pronunciation. This

process is straightforward for words that end in consonants, to which the suffix can attach directly in the same way it would to Dutch words. For English words that end in <e>, the formation can be less clear; for instance, analogy with words like *antennetje* and *anekdotetje* may lead speakers to spell the diminutive form of *website* as *websitetje*. However, the pronunciation should form the basis for the selection of a suffix, and the pronunciation ends in /t/, so the diminutive form is *websiteje* ("websiteje / websitetje", n.d.). The only word of this type that occurs in the corpus used for this thesis is *cakeje*.

Out of all anglicism tokens, 3.4% in the NL-OR corpus and 2.1% in the NL-TR corpus have a diminutive form. For about half of these occurrences, the diminutive suffix is attached to a Dutch element of the word (e.g. *mixdrankjes*, *jamkoekjes*). The table below only lists the occurrences in which the suffix is attached to an English segment, because this specifically shows the process of diminutising borrowed elements.

Table 4.4: Occurrences of English elements with a Dutch diminutive suffix

Loan	Occurrences in NL-OR	Occurrences in NL-TR
blendertje	1	0
burgertje (and variations)	7	0
cakeje	2	0
crackertje (and variations)	2	0
cupje	1	0
puddinkje (and variations)	0	2
dipje	1	0
toastje (and variations)	7	0
tripje (and variations)	2	0
total	**23** (2.5% of all anglicism tokens)	**2** (0.2% if all anglicism tokens)

These forms all correspond to the guidelines suggested by Onze Taal, so there does not seem to be a difference in conventionality between the two sub-corpora in this regard. The main difference here is in the frequency of the diminutives; 2.5% of anglicisms in the NL-OR corpus have diminutive suffixes attached to an English element compared to only 0.2% in the NL-TR corpus.

Verb conjugation, plural formation, and gender assignment

In addition to the advice discussed above, Onze Taal has articles on the spelling of English verbs in general, on pluralising English nouns in Dutch, on assigning a grammatical gender to English nouns, and on the conjugation of a list of specific

English verbs in Dutch. However, the analysis of these word classes and features revealed no significant differences in terms of conventionality between NL-OR and NL-TR. First, 18.8% of the NL-OR corpus and 8.5% of the NL-TR corpus are plural nouns with the plural suffix added to an English segment, and all but one of these (*chocoladelollies* in NL-OR) are formed according to Onze Taal's advice. Second, grammatical gender for all English nouns (or at least the words that have determiners or adjectives signalling their gender) is assigned as recommended; the majority are masculine, and the neuter words (5.3% in NL-OR and 2.6% in NL-TR) correspond to the exceptions outlined by Onze Taal. Finally, all of the verbs in the corpus are formed in correspondence with Onze Taal's guidelines, as well ("Adviezen over Engelse woorden in het Nederlands", n.d.).

List of common anglicisms

In addition to the advice on the integration of lexical items, Onze Taal provides a list of common words and expressions that have been translated directly from English (i.e. loan shifts). Most of these were not included in the initial anglicism count described in 4.2, because they are the result of structural borrowing or other processes that are not as easily identified as lexical borrowing. The number of occurrences of the terms from this list that occur in the cookbook corpus are listed in the table below, along with their degree of acceptability and the native equivalents suggested by Onze Taal.

Table 4.5: occurrences from Onze Taal's list of anglicisms in the NL-OR and NL-TR sub-corpora

Anglicism in Dutch	English original	Acceptable?	Native alternatives	Tokens in NL-OR	Tokens in NL-TR
grip	grip	yes	*greep*	1	0
huisgemaakt	homemade	yes	*eigengemaakt, handgemaakt, zelfgemaakt*	1	1
koffie maken	to make coffee	doubtful	*koffie zetten*	2	0
leven	to live (somewhere)	no	*wonen*	2	0
meer en meer	more and more	yes	*steeds meer, hoe langer hoe meer*	0	1
meer recentelijk	more recent(ly)	no	*recenter, onlangs nog, pas nog*	0	1
meest (e.g. "meest mooi")	most (beautiful)	sometimes	*mooist*	3	7
wereldwijd	worldwide	yes	*mondiaal*	4	1
wijd	wide	doubtful	*breed*	6	4
Total				**19**	**15**

Per 10,000 words, the NL-OR corpus contains 1.1 of these anglicisms that are considered "acceptable", 1.3 of which degree of the acceptability is doubtful, and 0.6 that are unacceptable. The NL-TR corpus, every 10,000 words contain 0.9 acceptable, 0.5 doubtful, and 0.5 unacceptable loan words. It should be noted that these numbers only apply to the anglicisms listed above—the corpora may contain others that Onze Taal has not yet listed. In any case, the numbers suggest that NL-TR contains fewer of these loan translations overall and that the loan translations it does contain are more likely to be acceptable. The table below shows the instances of these anglicisms as they occur in the corpus.

Table 4.6: Specific instances of anglicisms from Onze Taal's list as they occur in each sub-corpus

Anglicism	Occurrences in NL-OR	Occurrences in NL-TR
grip	(1) "grip te krijgen op kookprocessen" (Mariën & Groenewold).	0
huisgemaakt	(1) "huisgemaakte ricotta van geitenmelk" (Van Boven).	(1) "Huisgemaakte open taarten" (Hay).
koffie maken	(1) "hij maakte altijd een volle kan koffie toebroek"; (2) "hoe je lekkere koffie toebroek maakt" (Pereira, Van der Rijst & Stoel).	0
leven	(1) "wereldwijd verspreid levende bevolkingsgroepen" (Van der Velde & De Kroon); (2) "niet erg om in m'n uppie te leven" (Van Loon).	0
meer en meer	0	(1) "meer en meer Chinezen lijden aan obesitas" (Dunlop).
meer recentelijk	0	(1) "of, meer recentelijk, de speurtocht naar ingrediënten (...)" (Khoo).
meest	(1) "de meest pure olie" (Vreugdenhil); (2) "de meest onvergetelijke momenten" (Naessens); (3) "de meest gezonde vissen" (Naessens).	(1) "de meest exotische groente" (Elliot); (2) "het meest gangbare formaat" (Baugniet); (3) "de meest uiteenlopende vissoorten" (Bennett); (4) "het meest modieuze kledingstuk" (Khoo); (5) "de meest recente telling" (Lawson); (6) "de (...) meest exotische provisiekastmaaltijd" (Lawson); (7) "de meest effectieve manier" (Ottolenghi).
wereldwijd	(1) "wereldwijde bermsalade" (Bussink);	(1) "dat de problemen (...) wereldwijd nog net zo groot zijn als eerst" (Fearnly-

	(2) "wereldwijd verspreid" (Van der Velden & De Kroon); (3) "werk ik wereldwijd intensief samen" (Van Olphen); (4) "er wereldwijd mannen en vrouwen in hun keuken staan te zingen" (Bakker).	Whittingstall).
wijd	(1) "wijde, lage pan" (Vreugdenhil); (2) "wijde, lage schaal" (Luiten); (3) "wijde kommen" (Ten Houte de Lange & Van Lindonk); (4) "wijde kommen" (Ten Houte de Lange & Van Lindonk); (5) "wijde schaal" (Ten Houte de Lange & Van Lindonk); (6) "wijde braadpan" (Van der Velden & De Kroon).	(1) "wijde ovenvaste schaal" (Lawson); (2) "wijde pan" (Lawson); (3) "wijde pan" (Ottolenghi); (4) "wijde pan" (Fearnley-Whittingstall).

As stated in table 4.5, the use of *meest* to form superlatives is only acceptable in some cases, e.g. when the adjectives ends in certain phonemes or when the writer wants to add emphasis ("meest origineel / origineelst", n.d.). Both of the corpora contain cases where the use of "meest" is justified (e.g. *meest exotische*, where the adjective ends in <sch>) and cases where it seems unnecessary (e.g. *meest pure*). Out of the total number of superlatives in each corpus (which was extracted by searching for "*st" and "*ste", eliminating the irrelevant entries, and adding this number to the instances listed above), 1.6% in the NL-OR corpus and 3.5% of the NL-TR corpus use "meest" rather than the suffix <st>. Similarly, the NL-TR corpus contains more instances where *meer* is use to form comparatives. Considering the fact that *meest* is often used for added emphasis (as suggested by Onze Taal above), this difference may be an indication of translators' tendency to clarify.

4.5 Findings from the comparable corpus

The main purpose of including the source texts as an additional corpus was to examine the exact situations in which translators choose to borrow. Many of the anglicisms in the translated texts have been integrated into Dutch and accepted by most speakers so that the transfer of the source term may not even register as borrowing to most speakers; likely examples of this are *dressing*, *dipsaus*, and *tips*. Sometimes, the occurrence of English words is unavoidable because the author of the source text refers to their other books which do not have Dutch translations. In other cases, English terms may have been transferred consciously in order to preserve the local colour of the source text (e.g. *cockney-accent*), for their stylistic appeal (e.g. *mock mash* or *back-to-black spaghetti*), for precision (e.g. specific ingredients, brand names, or

cultivars such as *cheddar*, *Clearspring*, and *Granny Smith*), or simply because the English terms sound more appetising than native equivalents (e.g. *sticky* and *dumpling* vs. *plakkerig* and *deegbal*).

An interesting finding that resulted from the analysis of the EN-OR corpus is the fact that anglicisms in translation are not always derived from the source text itself, as the table below illustrates:

Table 4.7: English terms from the source texts that were translated with other anglicisms

Source text term	Target text term
all good	alles oké
baked beans	smoky bonen
black beans	kidneybonen
blend [noun]	mix
cool thing	coole trend
drive-through	drive-in
elevating	upgraden
fully signed-up vegetarian	hardcore vegetariër
griddle pan	grillpan
laid back	relaxte
liquidiser	blender
low-fat	light
on the high street	in supermarkten
pubs	bars
screw-top jar	jampot
set	'settelen'
shows	tv-serie
stick blender, immersion blender	staafmixer

The right-hand column shows the words as they occurred in the translated text. Without consulting the source text corpus for reference, one might assume that these words were transferred directly from the source text, but the left-hand column reveals a list of different source text items. Many of these translation choices were probably made because the target text terms are more familiar in Dutch; words like *jampot*, *supermarkt* and *mix* may not even be recognisably English to Dutch speakers. Other translations are more remarkable, like *smoky bonen* for *baked beans*; perhaps the original English term was considered too unfamiliar for Dutch readers, and *smoky* was simply selected as an appetising adjective for the title of the dish.

Chapter 5: Conclusion and discussion

5.1 Introduction

The purpose of this thesis has been to examine the presence of anglicisms in translated and original Dutch texts in order to determine whether Dutch writers and Dutch translators exhibit different tendencies with regards to their use of anglicisms. Because translators are said to tend towards explicitation and conventionalisation, the hypothesis was that the anglicisms in translated texts would be less frequent and more conventional. The corpus that formed the basis for this investigation consisted of two components, one made up of samples from original Dutch cookbooks (the NL-OR corpus) and one of samples from Dutch cookbooks that had been translated from English (the NL-TR corpus). For both of these corpora, all anglicisms were identified and analysed with the help of a corpus processing tool in order to identify any significant differences in frequency, function, grammatical category, and conventionality. This chapter will summarise the main findings of this analysis.

5.2 Main findings

Research questions

The main research questions for this thesis, as introduced in section 1.3, were as follows:
   (1) Do cookbooks that have been translated from English into Dutch contain
       fewer anglicisms than those that were originally written in Dutch?
   (2) Are there any differences between the anglicisms in translations and original
       texts in terms of type, function, and grammatical category?
   (3) Do translations contain anglicisms that are more conventional than those in
       non-translational text?

Frequency

In terms of frequency, the NL-OR corpus contained 1.4 times as many anglicisms as the NL-TR corpus, and the anglicisms in this corpus were selected from a wider variety of different terms (see tables 4.1 and 4.2). These numbers appear to confirm the suspicion that translated texts tend to be more resistant to anglicisms than Dutch original texts, and this result is consistent with Musacchio's finding that Italian translated texts tended to be lower in lexical borrowing from English than Italian original texts (Musacchio, 2005).

Types, grammatical categories, and functions

The analysis of different types, grammatical categories, and functions of anglicisms yielded proportions that were generally very similar for the NL-OR and NL-TR sub-corpora. In terms of the different types as described in sections 2.3, 3.4, and 4.3, the two relevant findings were that the NL-OR corpus contained more loan creations and the NL-TR corpus was richer in proper nouns. At face level, these findings could be seen as an indication of a higher degree of conventionality in translated text due to an avoidance of loan creations, but the difference simply seems to be caused by the difference in language of origin; when Dutch authors combine English elements to create titles (e.g. *24Kitchen*, *Yvestown*), these terms are categorised as loan creations, but when English authors do the same (e.g. *Deliaonline*), the transferred terms are categorised as loan words in the translation.

The classification based on grammatical category (see table 4.3) revealed proportions that correspond to Muysken's hierarchies (2010). The main difference here was that the NL-TR corpus contained more proper nouns such as brand and restaurant names relevant to the original author's location and titles of books with no Dutch translations. In terms of function, the main difference was that the NL-OR corpus contained more anglicisms that appeared to be chosen mainly for the prestige associated with English as a dominant language, while the anglicisms in the NL-TR corpus tended to serve the purpose of adding local colour.

As described in section 4.3, some of the anglicisms seem to have obtained a specific meaning in Dutch that differs from their source terms (e.g. *mix, cake*). This observation is consistent with Laviosa's finding that borrowings can "[wedge themselves] into an existing semantic field" to serve as a differentiator (Laviosa, 2007). Partly as a result of differentiations like these, anglicisms used in translation were sometimes found to correspond to different English words in the source text (see table 4.3 Another example is a translation of the term *loaf cake* in which the first segment was omitted because the second segment already carries its specific meaning in Dutch.

Conventionality

Consulting with two reference corpora showed no clear difference in conventionality between the two components of the cookbook corpus: the percentages were close together, with Görlach's *Dictionary of European Anglicisms* containing a slightly higher proportion of the anglicisms in the NL-TR corpus and the Corpus Hedendaags Nederlands yielding the opposite result (see section 4.4). A comparison of the sub-corpora in terms of adherence to Dutch spelling conventions provided more useful results. One of the findings was that the translated texts were more likely to contain additional hyphenation in compound formation, perhaps out of an attempt to make the terms easier to comprehend.

Structural borrowing was only examined briefly, but a few relevant differences came to the surface. The NL-OR corpus contained more occurrences of anglicisms from a list of loan shifts that are considered the result of "lazy translations" that are not generally seen as "correct". However, some of these specific patterns were more common in the NL-TR corpus than in NL-OR, such as the use of *meer* and *meest* to form comparatives and superlatives (see tables 4.5 and 4.6). The first finding is consistent with the theory that translated text tends to be more conventionalised than original text (see section 2.4), while the second finding corresponds to Toury's law of interference (1995), which notes that features from a dominant source language are likely to be transferred to the target language. The use of *meer* and *meest* may also be considered a form of explicitation due to the added emphasis.

5.3 Discussion

The results described above are mostly in line with the expectations set up by the theories and studies described in chapter 2: the translated texts contain fewer anglicisms in general, more cultural terms in the form of proper nouns, and more instances of explicitation due to added punctuation or emphasis. The tendency to conventionalise was not as apparent as expected; perhaps this universal feature of translation applies more to the native elements of the translation than to the transferred items. A surprising finding was the fact that in a significant number of cases, English terms from the source text were replaced by different English terms in the Dutch translation—although the corpus of source texts was not large enough to explore this phenomenon in much detail. These findings about translational behaviour are not necessarily of practical use to translators, but the fact that original texts proved to be so rich in anglicisms may indicate that translators do not need to be as resistant towards the use of transference as implied by Newmark (1980) (see section 2.2).

A problem with the classification of the functions of anglicisms depends on the interpretation of the reader, and the same anglicism may be categorised differently depending on which corpus it appears in. For instance, if an author of a trendy Dutch cookbook uses the word *soda bread*, this may be interpreted as partly motivated by the prestige of English as a fashionable language. When it is made by a translator, however, the decision to use the same borrowing may be attributed to a desire to add local colour by referring to items from the original author's culture. Clearly, the country of origin has at least some influence on the interpretation of the borrowings within any given text, and because of its richness in cultural terms and references, this true for food writing in particular. As a result, almost all translators' borrowings may be interpreted as attempts to stay close to the source text. However, the analysis of the corpus of source texts showed that translators do not necessarily limit themselves to the English words available in the source text, and when they

choose to replace source text terms with different anglicisms, they may be influenced by the same motivations that cause writers of original texts to use anglicisms.

The results of this thesis represent a very specific type of text and a specific language pair. The process of cookbook translation appears to be target-oriented and source-oriented at the same time: the representation of the source culture associated with a cookbook's author and cuisine are likely to have a foreignising effect, while adapting the recipe to be usable for inhabitants of the target language region is likely to involve domestication and adjustment to target language conventions. As a result, it does not seem reasonable to generalise the results found in this thesis to apply to other instances of translation. Nevertheless, the findings correspond to the results of Musacchio's study on English influence on translation (2005) and to some of the translation universals identified by Baker (1993) and Kruger (2002) (see section 2.4), so perhaps the results can be interpreted as additional evidence in support of their claims.

Bibliography

Adviezen over Engelse woorden in het Nederlands | Genootschap Onze Taal. (n.d.)
    Retrieved on 14 June, 2014, from
    https://onzetaal.nl/taaladvies/trefwoord/engels.

Anglicismen (leenvertalingen uit het Engels: juist of niet?) | Genootschap Onze Taal.
    (n.d.). Retrieved on 1 June, 2014, from
    https://onzetaal.nl/taaladvies/advies/anglicismen.

Baker, M. (1993). Corpus linguistics and translation studies: Implications and
    applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and
    technology: In honour of John Sinclair* (233-350). Amsterdam, the Netherlands:
    John Benjamins Publishing.

Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In
    H.L. Somers (Ed.), *Terminology, LSP, and translation: Studies in language
    engineering in honour of Juan C. Sager* (175-186). Amsterdam: John Benjamins
    Publishing.

Baker, M. (2007). Corpus-based translation studies in the academy. *Journal of foreign
    languages,* (5), 50-55.

Blum-Kulka, S. (1986). Shifts of cohesion and coherence in translation. In J. House
    & S. Blum-Kulka (Eds.), *Interlingual and intercultural communication: Discourse
    and cognition in translation and second language acquisition studies* (17-35).
    Tübingen, Germany: Gunter Narr Verlag.

bol.com | Bestverkochte kookboeken | Boeken. (2014). Retrieved on 29 April, 2014,
    from http://www.bol.com/nl/l/boeken/-
    /N/29/No/0/section/books/lijstid/31400001/index.html.

Bowker, L. & Pearson, J. (2002). *Working with specialized language: A practical guide to
    using corpora*. London, United Kingdom: Routledge.

Corpas Pastor, G. & Seghiri, M. (2007). Specialized corpora for translators: A
    quantitative method to determine representativeness. *Translation journal, 11*(3).
    Retrieved on 21 May, 2014, from
    http://www.translationdirectory.com/articles/article1445.php.

Corpas Pastor, G. & Seghiri, M. (2009). Virtual corpora as documentation resources:

Translating travel insurance documents. In A. Beeby, P. Rodríguez Inés, & P. Sánchez-Gijón (Eds.), *Corpus use and translating: Corpus use for learning to translate and learning corpus use to translate* (75-107). Amsterdam, the Netherlands: John Benjamins Publishing.

Corpus Hedendaags Nederlands. (2014). Retrieved on 14 June, 2014, from http://www.inl.nl/inl-nieuws/1012-corpus-hedendaags-nederlands.

Corpus Hedendaags Nederlands search. (2013). Retrieved on 14 June, 2014, from https://portal.clarin.inl.nl/shibsearch/page/search.

De Bestseller 60, de bestverkochte boeken wekelijks op een rij. (2014). Retrieved on 17 April, 2014 from http://www.cpnb.nl/bs/index.asp?gnr=15.

Engelse woorden in Nederlandse samenstellingen | Genootschap Onze Taal. (n.d.). Retrieved on 14 June, 2014, from https://onzetaal.nl/taaladvies/advies/engelse-woorden-in-nederlandse-samenstellingen.

Frankenberg-Garcia, A. (2005). A corpus-based study of loan words in original and translated texts. Retrieved on 24 January, 2014, from http://www.linguateca.pt/documentos/FrankenbergGarciaCL2005LoanWords.pdf.

Frankenberg-Garcia, A. (2009a). Are translations longer than source texts? A corpus-based study of explicitation. In A. Beeby, P. Rodríguez Inés, & P. Sánchez-Gijón (Eds.), *Corpus use and translating: Corpus use for learning to translate and learning corpus use to translate* (47-58). Amsterdam, the Netherlands: John Benjamins Publishing.

Frankenberg-Garcia, A. (2009b). Compiling and using a parallel corpus for research in translation. *Babel: International journal of translation, XXI*(1), 57-71. Retrieved on 21 May, 2014, from http://www.linguateca.pt/COMPARA/docum/FG2009CompilingUsingParallelCorpusForResearch.pdf.

Frawley, W. (1984). *Translation: Literary, linguistic, and philosophical perspectives*. Newark, Delaware: University of Delaware Press.

Galinsky, H. (1967). Stylistic aspects of linguistic borrowing: A stylistic view of American elements in modern German. In Carstensen, B. & Galinsky, H., *Amerikanismen der deutschen Gegenwartssprache*. Heidelberg, Germany: Winter.

Gerritsen, M. (1996). Engelstalige productadvertenties in Nederland: onbemind en

onbegrepen. In R. Van Hout & J. Kruijsen (Eds.), *Taalvariaties: Toonzettingen en modulaties op een thema* (67-83). Dordrecht, the Netherlands: Foris Publications.

Gerritsen, M., Korzilius, H., Van Meurs, F., & Gijsbers, I. (2000). English in Dutch commercials: Not understood and not appreciated. *Journal of advertising research, 40*(4), 17-36.

Gerritsen, M., Nickerson, C., Van Hooft, A., Van Meurs, F., Korzilius, H., Nederstigt, U., Starren, M., & Crijns, R. (2010). English in product advertisements in non-English-speaking countries in Western Europe: Product image and comprehension of the text. *Journal of global marketing*, 23, 1-17.

Görlach, M. (2003). *English words abroad*. Amsterdam, the Netherlands: John Benjamins Publishing.

Görlach, M. (2005). *A dictionary of European anglicisms*. New York City, New York: Oxford University Press.

Gottlieb, H. (2004). Danish echoes of English. *Nordic journal of English studies, 3*(2), 39-65.

Halverson, S. (1998). Translation studies and representative corpora: Establishing links between translation corpora, theoretical/descriptive categories and a conception of the object of study. *Meta: Translators' journal, 43*(4), 494-514.

Haspelmath, M. (2009). Lexical borrowing: Concepts and issues. In M. Haspelmath & U. Tadmor, *Loanwords in the world's languages: A comparative handbook* (35-53). Berlin, Germany: Walter de Gruyter.

Haugen, E. (1950). The analysis of linguistic borrowing. *Language, 26*(2), 210-231.

Hilgendorf, S.K. (1996). The impact of English in Germany. *English Today, 12*(3), 3-14.

Kruger, A. (2002). Corpus-based translation research: Its developments and implications for general, literary and Bible translation. *Acta Theologica, Supplementum 2, 22*(1), 70-106.

Laviosa. S. (2002). *Corpus-based translation studies: Theory, findings, applications*. Amsterdam, the Netherlands: Rodopi.

Laviosa, S. (2007). Studying anglicisms with comparable and parallel corpora. *Belgian Journal of Linguistics, 21*(1), 123-136.

meest origineel / origineelst | Genootschap Onze Taal. (n.d.). Retrieved on 14 June, 2014, from https://onzetaal.nl/taaladvies/advies/meest-origineel-origineelst.

Musacchio, M.T. (2005). The influence of English on Italian: The case of translations of economics articles. In G. Anderman and M. Rogers (Eds.), *In and out of English: For better, for worse?* (71-96). Clevedon, United Kingdom: Multilingual Matters.

Muysken, P. (2010). Scenarios for language contact. In R. Hickey (Ed.), *The handbook of language contact* (265-281). Chichester, United Kingdom: John Wiley & Sons.

Myers-Scott, C. (2006). *Multiple voices: An introduction to bilingualism*. Oxford, United Kingdom: Blackwell Publishing. pp. 212-118.

Nesselhauf, N. (2007). Diachronic analysis with the internet? Will and shall in ARCHER and in a corpus of e-texts from the web. In M. Hundt, N. Nesselhaulf, & C. Biewer (Eds.), *Corpus linguistics and the web* (287-305). Amsterdam, the Netherlands: Rodopi.

Newmark, P. (1988). *A textbook of translation*. Hemel Hempstead, United Kingdom: Prentice Hall.

Onysko, A. (2004). Anglicisms in German: from iniquitous to ubiquitous?. *English Today*, *20*(1), 59-64.

Onysko, A. & Winter-Froemel, E. (2011). Necessary loans – luxury loans? Exploring the pragmatic dimension of borrowing. *Journal of pragmatics, 43*(6), 1550-1567.

Ridder, S. (1995). Dutch in English. *English today, 11*(4), 44-50.

Toury, G. (1995). *Descriptive translation studies and beyond*. Amsterdam, the Netherlands: John Benjamins Publishing.

Tymoczko, M. (1998). Computerized corpora and the future of translation studies. *Meta: Translator's journal, 43*(4), 652-660.

Van der Sijs, N. (2010). Etymologiebank. Retrieved on 1 June, 2014, from http://www.etymologiebank.nl.

Van der Sijs, N. (2012). Engelse leenwoorden revisited: Hoeveel wordt het Nederlands gemixt met het Engels?. *Onze Taal*, 5, (132-133). Retrieved on 14 June, 2014, from https://onzetaal.nl/uploads/editor/leenwoordtelling.pdf.

Van Hout, R. & Muysken, P. (1994). Modelling lexical borrowability. *Language variation and change*, (6), 39-62.

Venuti, L. (2000). *The translation studies reader*. London, United Kingdom: Routledge.

Vinay, J.P., & Darbelnet, J. (2000). A methodology for translation. (J.C. Sager & M.-J. Hamel, Trans.). In L. Venuti, *The translation studies reader* (84-93). London, United Kingdom: Routledge.

websiteje / websitetje | Genootschap Onze Taal. (n.d.). Retrieved on 14 June, 2014, from https://onzetaal.nl/taaladvies/advies/websiteje-websitetje.

Zanettin, F. (2002). CEXI: Designing and English Italian translational corpus. *Language and computers*, 42(1), 329-343.

Appendix 1A: Books used for the NL-OR corpus

Book title	Authors	Publisher	Year	Number of words in excerpt
70 Groene Smoothies	Marjolijn van der Velde	De Groene Uitgever	2013	721
Boekoe Bangsa	Harold Pereira. Mirjam van der Rijst. and Eveliene Stoel	Terra	2011	848
Comfort Food	Janneke Vreugdenhil	Nieuw Amsterdam	2011	2349
De Dunne Vegetariër	Antoinette Hertsenberg	Karakter	2009	1684
DedikkevanDam	Johannes van Dam	Nijgh & Van Ditmar	2012	1909
Down to Earth	Sacha de Boer and Jacob Jan Boerma	Van Dishoeck	2013	2178
Ellemieke's Familie Kookboek	Ellemieke Vermolen and Annelies Rutten	Minestrone	2013	2112
Eten uit de Natuur	Michiel Bussink	Van Duuren Media	2013	1792
Green Delicious	Natascha Boudewijn	Becht Lifestyle	2012	1165
Grenzeloos Koken		Karakter	2012	1788
Het Beste Dieet van de Wereld	Christian Bitz and Arne Astrup	Kosmos	2013	534
Home Made Zomer	Yvette van Boven	Fontaine	2013	929
Impress Your Friends	Angélique Schmeink	Karakter	2014	3937
Koken met Kennis	Eke Mariën and Jan Groenewold	Karakter	2013	1490
Lekker Lang Jong	Clara ten Houte de Lange and Nelleke van Lindonk	Lindonk & de Bres	2013	5144
Oergondisch Genieten	Ria Penders and Yvonne van Stigt	Oergezond	2014	2591
Puur Genieten 2 En Toch Gezond en Slank	Pascale Naessens	Lannoo	2013	4070
Rudolph's Cupcakes Compleet	Rudolph van Veen	Karakter	2013	1299
Rutger Bakt	Rutger van den Broek	Carrera	2014	2155
Smart Cooking Compleet	Julius Jaspers	Carrera	2011	3443
Superfood Recepten	Jesse van der Velde and Annemieke de Kroon	Spectrum	2013	2904
The Culy Way of Life	Monique van Loon	Bertram+De Leeuw Uitgevers	2013	1347
Werken met Vis	Bart van Olphen	Carrera	2012	1804
Winterslank	Sonja Bakker	De Zonnestraal	2011	4940
Yoga Kookboek	Kyra de Vreeze	Fontaine	2013	1432
Yvestown in de Keuken	Yvonne Eijkenduijn	Snor	2014	573
Koken met Karin Zónder Pakjes en Zakjes / 2	Karin Luiten	Nieuw Amsterdam	2013	8031
Total number of words				**63169**

Appendix 1B: Books used for the NL-TR corpus

Book title	Authors	Original title	Translators	Publisher	Year	Number of words in excerpt
(H)eerlijk	Gwyneth Paltrow	It's all good	Ingrid Buthod-Girard, Akkie de Jong	Kosmos	2013	754
30 Minuten Vegetarisch	Rose Elliot	30-minute vegetarian	Aniek Njiokiktjien	Kosmos	2012	2149
365 Sappen & Smoothies	Natalie Savona	The big book of juices and smoothies	André Kaijim	Veltman	2005	1413
500 Stoof- & Ovenschotels	Rebecca Baugniet	500 casseroles	Marijne Thomas	Veltman	2012	6180
500 Sushi	Caroline Bennett	500 sushi dishes	Anna Penta	Veltman	2013	637
500 Taarten en Cakes	Susannah Blake	500 cakes	Ewout Hanselaar	Veltman	2011	815
Baking Made Easy	Lorraine Pascale	Baking made easy	Ireen Niessen	Veltman	2012	1418
Bakken met de Cake Boss	Buddy Valastro	Baking with the Cake Boss	Maaike van der Rijst	Veltman	2013	600
Chez Rachel	Rachel Khoo	The little Paris kitchen	Vitataal	Kosmos	2012	5398
De Echte Chinese Keuken Thuis	Fuchsia Dunlop	Every grain of rice	Jacqueline IJsselstijn, Ewout Hanselaar	Karakter	2013	1736
De Free Range Cook	Annabel Langbein	The free range cook	Jolanda Abbes, Ron ter Borg	Van Dishoeck	2012	1337
De Kunst van het Bakken	Delia Smith	Delia's cakes	Ester van Buuren	Terra	2013	1209
De Smaken van Spanje	Claudia Roden	The food of Spain	Jacques Meerman	Fontaine	2013	1061
De Souk	Salma Hage	The Lebanese kitchen	Henja Schneider, Jaro Schneider	Van Dishoeck	2013	2486
Delicious. Lekker Koken!	Valli Little	Delicious. simply the best	Hennie Franssen-Seebregts	Fontaine	2012	1791
De Basics	Donna Hay	A cook's guide	Ester van Buuren	Van Dishoeck	2014	7636
Heston Blumenthal Thuis	Heston Blumenthal	Heston at home	Saskia Peeters, Mariëlle Steinpatz, Dido Tchaoussoglou	Karakter	2012	2447
Het Grote Granenboek	Ghillie James	Amazing grains	Jacques Meerman	Fontaine	2014	2482
Italiaanse Hapjes	Lindy Wildsmith and Valentina Harris	Cicchetti: and other small Italian dishes to share	Ellen Hosmar	Veltman	2014	808
Jamie in 15 Minuten	Jamie Oliver	Jamie's 15 minute meals	Jaromir Schneider	Kosmos	2012	8526
Koken met Quinoa	Rena Patten	Cooking with quinoa	Mariëlle Steinpatz	Veltman	2013	677
Matt Preston's 100 Beste Recepten	Matt Preston	Matt Preston's 100 best recipes	Félice Portier	Kosmos	2013	833
Nigellissima	Nigella Lawson	Nigellissima	Henja Schneider	Atlas Contact	2012	7879
Plenty	Yotam Ottolenghi	Plenty	Hennie Franssen-Seebregts	Fontaine	2013	10355

Raw Food als Levenskunst	Jenny Ross	The art of living food	Engelien Scholtes	Koppenhol	2009	432
Roken, Drogen en Pekelen	Dick Strawbridge and James Strawbridge	Curing & smoking	Claudia Dispa	Fontaine	2013	1014
Veg!	Hugh Fearnley-Whittingstall	River Cottage veg everyday!	Roselle de Jong	Becht	2012	3277
Total number of words						**75350**

Appendix 2: List of anglicisms

Loan word	OR occurrences	TR occurrences	Variations
24Kitchen	2		
ale		2	
allspice	1		
artwork	1		
As for butter versus margarine, I trust cows more than chemists	1		
Aspro-Clear		1	
baby	2	1	baby (TR: 1), babyoctopus (OR: 2)
back-to-black		1	
bacon		14	bacon (TR: 13), baconplakken (TR: 1)
bagel		1	
banana joy	1		
barbecue	8	8	BBQ (OR: 4), bbq (OR: 1), barbecue (OR: 3, TR: 3), barbecuesmaak (TR: 1), barbecuesaus (TR: 4)
bars		1	
beat		1	
beef	14	20	beef (TR: 1), biefsalade (TR: 1), biefstuk (OR: 7, TR: 3), biefstukje (OR: 1), biefstukjes (OR: 2, TR: 4), biefstukken (OR: 2, TR: 5), biefstukreepjes (OR: 1), herten- (of runder)biefstukken (TR: 1) lamsbiefstukjes (TR: 3), (lende)biefstukken (TR: 1), lendebiefstukjes (TR: 1), rosbief (OR: 1)
bestseller	1		
bite	2		
blender	50	20	blend (OR: 18), blender (OR: 28, TR: 20), blenders (OR: 1) blendertje (OR: 1), personal blender (OR: 2)
blog	1		
blue stilton	3		
boogie	1		
boost	3		boost (OR: 2), energieboost (OR: 1)
Bounty	2		
brainstorm	1		
brandy		3	
brownies		1	
bubble and squeak		2	
budget	1	1	
burger	7	6	burger (TR: 2), burgers (TR: 2), burgertje (OR: 2), burgertjes (OR: 2), gehaktburgertjes (OR: 1), courgette-gehaktburgertjes (OR: 1), hamburger (TR: 1), hamburgers (TR: 1) hamburgertje (OR: 1)
cacaonibs	1	5	
caesar salad	2		
cajun		4	cajungarnalen (TR: 1), cajunkruiden (TR: 2), cajunsteak (TR: 1)
cake	19	23	appelcake (OR: 2), basiscake (TR: 1) cake (OR: 11, TR: 11), cakeje (OR: 2), cakelagen (TR: 1), cakerecept (TR: 1), cakes (OR: 2, TR: 7), chocoladecake (OR: 1), chocoladecakes (OR: 1) citrusvruchtencake (TR: 1), plaatcake (TR: 1)
cashewnoten	19		
cater	2		cateraar (OR: 1), cateringbedrijf (OR: 1)
cheddar		8	
cheesecake		1	
chemist	1		
cherry pie	1		

cherrytomaten	7	3	cherrytomaten (OR: 6, TR: 1), cherrytomaatjes (OR: 1, TR: 2)
chick weed	1		
chicken sticks	1		
chief	1		
chillen	1		
chips	6	6	aardappelchips (OR: 3), cassave chips (TR: 1) chips (OR: 1, TR: 1), tortillachips (TR: 4), pastinaak chips (OR: 1), schorsenerenchips (OR: 1)
chocoholic		1	
chocolate chip cookies	1		
chocolate chips		1	
chutney	3	1	chutney (TR: 1), mangochutney (OR: 2), mango chutney (OR: 1)
Clearspring		2	
cockney-accent		1	
cocktail	5	2	cocktails (OR: 2), garnalencocktail (OR: 2, TR: 1), whisky-cocktailsaus (OR: 1) wodkacocktails (TR: 1)
cole slaw		1	
comfort food	1	2	
comfortkoken	1		
computer	1		
cook	1		
Cook & Chemist	3		
coole		1	
Countryachtige	1		
cracker	7		crackers (OR: 3), sesamcrackers (OR: 1), lijnzaadcrackers (OR: 1), crackertje (OR: 1), crackertjes (OR: 1)
cranberry's	1		
cream cheesecrème	4		aardbeien-cream cheesecrème (OR: 3), cream cheesecrème (OR: 1)
cream sherry		1	
cress		6	cressen (TR: 3), koriandercress (TR: 3)
crisp	1		
Crown Prince		1	
crumble	9		crumble (OR: 3), hazelnootcrumble (OR: 3), pecancrumble (OR: 3)
crunch		1	
cumberlandworstjes		1	
cup	4	2	courgettecups (OR: 3), cup (TR: 1), cupje (OR: 1), cupmaten (TR: 1)
cupcake	43	1	cupcake (OR: 5), cupcake-pavlova (TR: 1) cupcakes (OR: 29) cupcakeblik (OR: 1), cupcakevormpjes (OR: 4), cupcakemix (OR: 1), cupcake-ijsjes (OR: 1), cupcaketray (OR: 2)
curd	9		curd (OR: 2), mango curd (OR: 1), mangocurd (OR: 6)
curry	15	21	curry (OR: 4, TR: 9), curryblaadjes (TR: 4), currypasta (OR: 6, TR: 5), currypasta's (OR: 1, TR: 1) currypoeder (OR: 2), dahlcurry (TR: 1), groentecurry (OR: 1, TR: 1), groentecurry's (OR: 1)
custard	2	7	custard (OR: 1, TR: 5), custardpoeder (OR: 1), custardpuddinkjes (TR: 1), vanillecustard (TR: 1)
De Basics		1	
dehydratorlade		2	
Deliaonline		1	
diepvries	19	16	diepvries (OR: 12, TR: 7), diepvriesbak (OR: 2), diepvriesdeeg (TR: 1), diepvriesdoperwten (OR: 1, TR: 5), diepvriesdoos (TR: 1), diepvriesgroenten (OR: 1), diepvriesmaïskorrels (TR: 1), diepvriestuinbonen (TR: 1), diepvriezer (OR: 3)
digestives		1	

dinner for one	2		
dip	13	6	dip (OR: 5, TR: 2), dipje (OR: 1), dippen (OR: 2), dips (OR: 1), dipsaus (TR: 4), doperwtendip (OR: 2), korianderdip (OR: 2)
donuts		2	
double cream		1	
down to earth	5		Down to Earth (OR: 4), down to earth (OR: 1)
dressing	30	35	basisdressing (OR: 2), dressing (OR: 22, TR: 27), dressings (OR: 2, TR: 2), karnemelkdressing (TR: 2), knoflookdressing (TR: 1), kruidendressing (OR: 1), mosterddressing (TR: 1), signatuurdressing (TR: 1), sinaasappeldressing (OR: 3), sojadressing (TR: 1)
drive-in		1	
drumsticks		2	
dry	4		
dumplings		2	
east meets west		1	
e-book	1		
eggs Benedict	1		
essay		1	
ever	1		
evergreens	1		
extra large		1	
fan	2		
fashionable	1		
fastfood		3	
Fat Duck		3	Fat Duck (TR: 2), Fat Duckgerechten (TR: 1)
feelgoodaroma	1		
fish & chips	1		
fitheid	1		
flakes	2		zoutflakes (OR: 1), zeezoutflakes (OR: 1)
flatbread		5	flatbread (TR: 4), flatbreaddeeg (TR: 1)
Food	1		
foodie	1		
foodprocessor	3	7	
Fresh from the Freezer		1	
fulltime	1		
funky	2		
fusion		1	
gel	4		aardbeiengel (OR: 2), gel (OR: 2)
ghee	2	2	
gin		1	
golden syrup		2	
goodies	1		
Google	1		
granny smith	1	3	Granny Smith (OR: 1, TR: 2), granny smith-appel (TR: 1)
Granny Takes A Trip		1	
granola		3	granola (TR: 1), ontbijtgranola (TR: 1), quinoa-granola (TR: 1)
grapefruit	6	15	grapefruit (OR: 3, TR: 10), grapefruits (OR: 2, TR: 3), grapefruitplakken (OR: 1), grapefruit-vitaliteit (TR: 1), grapezoet (TR: 1)
grill	26	62	gegrild (TR: 1), gegrilde (OR: 6, TR: 13), gril (OR: 1, TR: 4), grill (OR: 10, TR: 20), grillen (OR: 2, TR: 3), grillmarkeringen (TR: 1), grillpan (OR: 4, TR: 16), grillplaat (OR: 2), grillstand (OR: 1), ovengrill (TR: 4)
grip	1		

groovy		1	
groupie		1	
Guardian		7	Guardian (TR: 5), Guardian-lezer (TR: 1), Guardianlezers (TR:1)
Halloween		4	Halloween (TR: 1), halloweenpakketten (TR: 1), halloweensoufflés (TR: 2)
Happy Cooking!		1	
hardcore		1	
heartbreaking	1		
high tea	1		
hip	2		
History of Western Philosophy		1	
hittefreaks		1	
hobbit	1		
hobby	2		hobbyist (OR: 1), hobbykok (OR: 1)
home alone	1		
home sweet home	1		
homemade	1		
hotspots	1		
how-to	2		how-to (OR: 1), how-to's (OR: 1)
How-to-Cook-colums		1	
HP sauce		1	
i scream	1		
icing	9		
impact	1	1	
Impress Your Friends	1		
Indian	1		
ingesprayde	1		
Instagram	1		
instant	2	5	instant (OR: 2), instant-espressokoffie (TR: 2), instantkoffie (TR: 3)
interactie	1		
internet	1		
Irish		1	
jam	12	26	aardbeienjam (OR: 2, TR: 1), abrikozenjam (OR: 2), chili-jam (TR: 11), frambozenjam (TR: 4), jam (OR: 1, TR: 8), jamkoekjes (TR: 1), jampot (TR: 1), kersenjam (OR: 7)
Jam, Jelly and Relish		1	
jambalaya		4	
jazzmusici		1	
jelly bean	2		jelly bean (OR: 1), jelly beans (OR: 1)
jerk		4	jerk-ingrediënten (TR: 1), jerk-saus (TR: 1), jerk-varkensfilet (TR: 2)
jetset	1		
jiggy jiggy-groenten		2	
jonagold	1		
junkfood		1	
ketchup	5	4	ketchup (OR: 4, TR: 1), tomatenketchup (OR: 1, TR: 3)
kick		2	
kidneybonen		3	
kids	2		
King's Road		1	
kiwi	8		kiwi (OR: 2), kiwi's (OR: 3), kiwipuree (OR: 2), kiwisap (OR: 1)
lamsrack	6		

Launceton Place		1	
Leiths School		1	
lemon	1		
lemon curd	1	2	lemon curd (TR: 2), lemoncurd (OR: 1)
lifestyle	2		lifestyle (OR: 1), lifestyletijdschriften (OR: 1)
light	9	4	light (OR: 9), light-kokosmelk (TR: 4)
little gem	5	5	little gem (OR: 5, TR: 1), littlegemsla (TR: 4)
live	1		
logo's	1		
lolly	3		lolly's (OR: 1), chocoladelolly (OR: 1), chocoladelollies (OR: 1),
lunch	13	13	gezinslunch (TR: 1), lievelingslunches (OR: 1), lunch (OR: 8, TR: 7), lunchen (TR: 1), lunches (TR: 1), lunchgerecht (OR: 2, TR: 2), lunchpakket (OR: 1), lunchtijd (TR: 1), meeneemlunch (OR: 1)
M&M's	2		
magazine	1	3	
Maldon		3	Maldon-zout (TR: 1), maldon zeezout (TR: 2)
Marine Stewardship Council	1		
marshmallow	2		marshmallow (OR: 1), marshmellows (OR: 1)
match made in heaven		1	
medium		8	medium (TR: 3), medium rare (TR: 2), medium-rare (TR: 1), mediumrare (TR: 1), medium-well (TR: 1)
mintjelly	1		
mix	77	46	broodkruimelmix (TR: 1), chocolademix (OR: 1), cupcakemix (OR: 1), gemixt (OR: 2), groentemix (OR: 2), handmixer (OR: 4, TR: 3), kruidenmix (OR: 2), mix (OR: 30, TR: 9), mixdrankjes (TR: 1), mixen (OR: 1, TR: 1), mixer (OR: 10, TR: 3), mixt (TR: 2), ochtendmix (TR: 1), oliebollenmix (OR: 2), paneermix (OR: 1), sinaasappelmix (TR: 1), staafmixer (OR: 18, TR: 14), (staaf)mixer (OR: 1), tafelmixer (TR: 1), venkelmix (OR: 2), zoutmix (TR: 9)
mix en match		1	
muffin	19	14	appelmuffins (OR: 1), bananenmuffins (TR: 1), briochemuffins (TR: 1) chiamuffins (OR: 2), havermuffins (OR: 1), muffin (TR: 1), muffinbakjes (TR: 1), muffinplaat (TR: 1), muffins (OR: 10, TR: 1), muffiningrediënten (OR: 1), muffinvorm (TR: 5), muffinvormpjes (OR: 3, TR: 2), sinaasappelbloesemmaanzaadmuffins (OR: 1), sinaasappelbriochemuffins (TR: 1)
Not Just a Load of Lentils		1	
nuggets		5	nuggets (TR: 2), kipnuggets (TR: 3)
office-mesje	1		
oké	1	2	
Old Man Mike		1	
online	2	4	
overall	1		
pancakes	3		
paperback		2	paperbackrechten (TR: 1), paperbackuitgever (TR: 1)
peaceful	1		
picknick		3	picknick (TR: 1), picnicken (TR: 1), picknicks (TR: 1)
pie		4	broccolipie (TR: 1), pie (TR: 3)
pingpongbal	1		
plant	1		
plastic	17	12	plastic (OR: 13, TR: 4), plasticfolie (OR: 4, TR: 8)
pop- en rockcultuur	1		
pop-upstudiootje	1		
pound of bananas, a		1	

pound!			
power	2	2	power (OR: 2), powerhavermout (TR: 1), powerontbijt (TR: 1)
pudding		12	custardpuddinkjes (TR: 1), griesmeelpudding (TR: 3), karamelpuddinkjes (TR: 1), pudding (TR: 6), rijstpudding (TR: 1)
puffy	1		
pulse		8	pulseknop, (TR: 7), pulsen (TR: 1)
quorn	1		
rare		5	medium rare (TR: 2), mediumrare (TR: 1), medium-rare (TR: 1), rare (TR: 1)
recoveryverpleegkundige	1		
red summer	1		
relaxte	1	1	
relish		1	
ribeye	4	1	ribeye (OR: 1, TR: 1), ribeyes (OR: 3)
River Cottage		1	
roast	1		
Rolling Stones		1	
Rolls-Royce	1		
rosary		1	
rub	1		
runt	2		
Sainsbury's		1	
Sainsbury's Magazine		1	
sandwich		4	sandwich (TR: 3), soezensandwiches (TR: 1)
scones	1	8	
scoren	1	1	scoren (TR: 1), scoort (OR: 1)
scotch bonnet		1	
Scotch eggs		2	
secret	1		
settelen		1	
sexy		1	
shepherd's pie		1	
sherry	2	12	manzanillasherry (TR: 1), sherry (OR: 2, TR: 11)
shortcuts	1		
silken	2		
Silverwood		1	
Simply Delicious		2	
single cream		1	
slogan		1	
slow cooking		1	
slow food		3	
slow start	1		
smartie	6		smartie (OR: 1), smarties (OR: 5)
smoky		2	
smoothie	21	3	smoothie (OR: 13, TR: 1), smoothies (OR: 7, TR: 2), smoothie-boek (OR: 1)
snack		4	middagsnack (TR: 1), snack (TR: 2), snacktijd (TR: 1)
so this is Christmas		1	
soda bread	1		
South Coast Farms		1	
spicy	1		
spread	4	1	groentespread (TR: 1), spreads (OR: 1), vruchtenspread (OR: 3)

spring roll pastry	2		
spring roll skin	1		
sprinkles		3	sprinkles (TR: 2), fetasprinkles (TR: 1)
stap-voor-stapformat		1	
steak	3	11	cajunsteak (TR: 1), steak (TR: 2), steaks (OR: 1, TR: 1), tonijnsteaks (OR: 2, TR: 7)
stereo	1		
stew		3	
sticky		1	
Stilton		1	
stylen	1		
sugarsnaps		2	
superfood	45		superfood (OR: 9), superfoodpatat (OR: 1), superfoods (OR: 35)
supermarkt	19	14	supermarkt (OR: 17, TR: 5), supermarkten (OR: 2, TR: 7), supermarktketen (TR: 1), supermarktketens (TR: 1)
Swiss Roll		1	
tabasco	2	1	
tattoo's	1		
tea parties		1	
team		3	
tearjerkers	1		
teddybeer	1		
test	2	9	getest (TR: 1), test (TR: 4), testen (OR: 1, TR: 3), testfoto's (OR: 1), testte (TR: 1)
The Great British Bake Off	1		
timing		2	
tip	90	26	tip (OR: 63, TR: 11), tips (OR: 25, TR: 15), combinatietip (OR: 1), wijntips (OR: 1)
to biologisch or not to biologisch	1		
toad-in-the-hole		2	toad-in-the-hole (TR: 1), toads-in-the-hole (TR: 1)
toast	7	5	avocadotoast (TR: 1), toast (TR: 3), toastje (OR: 4), toastjes (OR: 3), toastkruim (TR: 1)
toffees	2	1	karamel-toffeesaus (OR: 1), toffees (TR: 1), toffeesaus (OR: 1)
topping	10	9	chocoladepastatopping (OR: 3), pizzatopping (TR: 1), topping (OR: 5, TR: 8), toppings (OR: 2)
trend	3	1	
trip	3		trip (OR: 1), tripje (OR: 1), (pers)tripjes (OR: 1)
trolley	1		
tv	3	1	televisie (OR: 1), tv-programma (OR: 2), tv-serie (TR: 1)
twilight zone	1		
underground		1	
Union Square		1	
upgraden		1	
up-to-date	2		up to date (OR: 1), up-to-date (OR: 1)
vegan		16	veganist (TR: 2), veganisten (TR: 1), veganistisch (TR: 9), veganistische (TR: 4)
Vegetarian Mother and Baby Cook		1	
wc's		1	
website	3		website (OR: 2), websites (OR: 1)
weekend	2	2	
well done		3	medium-well (TR: 1), well done (TR: 2)
Welsh rarebit		1	
whisky	3	1	whisky (OR: 1), whisky-cocktailsaus (OR: 1), whiskyijs (TR: 1), whiskyvaten (OR: 1)

woodchips	2		
woolsery		1	
worcestershiresaus	3	9	worcestersaus (OR: 3, TR: 5), worcestershiresaus (TR: 4)
workshop	3	1	(kook)workshops (OR: 1), workshops (OR: 2, TR: 1)
wrap	10		koolwrap (OR: 1), kerriekokoswraps (OR: 1), kippenwraps (OR: 2), tortillawraps (OR: 2), wrap (OR: 1), wraps (OR: 3)
yum	1		
yvestown	1		
Total (tokens)	936	803	
Total (types)	190	180	