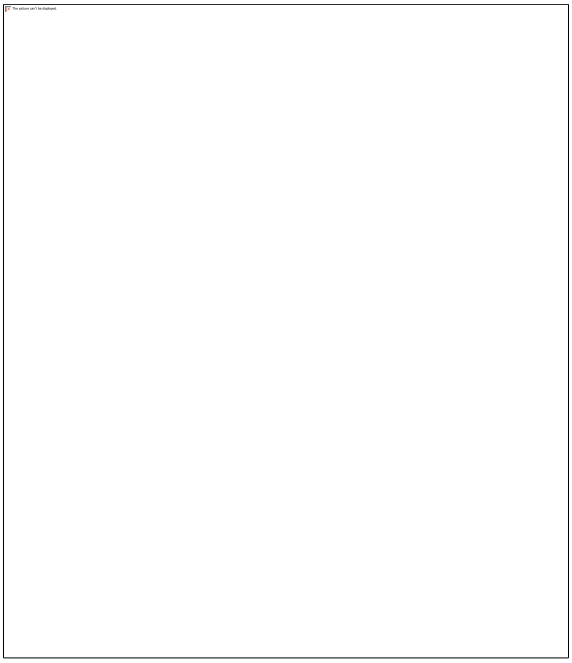
Dutch loanwords in American English: An increase in frequency or a decrease? A corpus study

ter Weer, Izra

Dutch loanwords in American English: An increase in frequency or a decrease?

A corpus study



By Izra A.M.H. ter Weer

S1773658

MA Thesis Linguistics: Modern Languages, Faculty of Humanities, Leiden University

Thesis Supervisor: Dr M.S. Gordon

Second Reader: Dr D. Smakman

Date of Submission: 21.06.2022

ACKNOWLEDGEMENTS

Writing and finishing this master's thesis has felt like finishing a chapter in the book about my life. Hopefully, completing this thesis will complete my student life: a period of six years full of joy, happiness, learning and studying. Firstly, I would like to deeply thank my thesis supervisor, Mo, for her feedback, support and quick responses. Even though she had multiple other students to supervise, when speaking to her, she always had her full attention on me. Next, I would like to thank my family for their genuine love, support and interest. Specifically, my mother for her advice regarding the English language, and my father for making sure I actually write my thesis, rather than watch Netflix. Lastly, I thank my friends for taking my mind off this thesis when necessary and for making sure I had enough fun events in between. This thesis is dedicated to you.

ABSTRACT

When the Dutch settled in North America during the seventeenth century, for colonising and trade purposes, came language contact between the two groups of speakers. One result from this contact is loanwords. This research investigates Dutch loanwords in American English and their frequency. Through a combination of Schultz's (2012, 2017, 2018, 2019) framework, the *Oxford English Dictionary* and Van der Sijs' (2010) glossary of Dutch loanwords in North American languages, this research was performed. The frequencies were analysed through the *Corpus of Historical American English* and the *Corpus of Contemporary American English*. These corpora provided an overview of the development of frequencies both over time and across text types. This research concludes that the frequency of Dutch loanwords in American English increases over time.

Keywords*: Loanwords, Dutch, American English,* OED, COHA, COCA, *frequency*

TABLE OF CONTENT

1 Introduction

"More words per capita have been borrowed into American English from [the] early Hollanders than from any other sort of non-English speakers" (Laird, 1970, p. 314).

Borrowed words, the result of the process of borrowing, are also known as loanwords. These words are borrowed from one language, the source language, and (permanently) adopted by the speakers of a different language. Loanwords have been around since the first occurrences of language contact. One of these instances of language contact is that between Dutch and American English. When Dutch immigrants arrived to the East Coast of America, in the seventeenth century, the first incidences of language contact between Dutch speakers and American English speakers took place. Whilst there is language contact between Dutch speakers and (American) English speakers today too, it is different from the historical situation. The former is a more casual form of language contact, whilst the latter is more intense on a settlement basis. Although language contact between Dutch speakers and American English speakers lasted relatively short, approximately 40 years, Dutch and its influence on American English proceeded for a notably longer time, due to several factors, such as the ongoing use of Dutch in churches.

The processes of borrowing and the adoption of loanwords as a result of one language influencing another have been studied by some researchers (Poplack, Sankoff & Miller 1988; Treffers-Daller 2010; Schultz 2012, 2017, 2018, 2019; to name a few). However, the amount of research that has been done on this topic, which is rather limited, does not reflect the strong presence of loanwords in the lexical inventories of languages. Likewise, the contact between Dutch and American English and the consequent loanwords into American English has been researched (Llewellyn 1936; van der Sijs 1998, 2010), but have not been studied extensively, so there is a niche here to fill. What is missing, amongst other things, is a corpus analysis investigating the frequency of loanwords. Additionally, there might still be questions about the social meaning of loanwords. Looking at loanwords in corpora can help determining analyses about social attitudes.

A researcher who has done extensive research on the impact of foreign languages on the English language is Schultz (2012, 2017, 2018, 2019). This author has researched the influence of French, Japanese, German, Spanish, and Yiddish on English. Schultz's theoretical considerations allow the author to investigate which borrowings from a foreign language, for example French, have entered the English language during a specific time period, for example the twentieth century. The approach Schultz uses is useful to study loanwords in general as it illustrates an up-to-date report of a fundamental proportion of the foreign vocabulary entering English. Not only does this author's framework trace the etymologies, it also considers other factors, such as the many spheres and areas of life in the English lexicon influenced by foreign languages. Schultz has not, however, researched the influence of Dutch on American English. This thesis will conduct this research using Schultz's framework. The author's framework utilises the advanced search-function of the *OED Online*. By entering the desired language, for example French, in the etymology of the words, the search results in loanwords from, for example, French in English and their consequent categories, i.e., fields of life. When applying Schultz's framework to Dutch, the expected outcome is a list of Dutch loanwords in English and resultant categories.

Van der Sijs (2010) has done thorough research on the long-lasting impact of the Dutch language in North America in her book *Cookies, Coleslaw, and Stoops*. The title of this book already gives three Dutch loanwords in American English away. The Dutch title of this book reveals another two loanwords: *Yankees, cookies en Dollars*. Whilst Van der Sijs (2010) has performed exhaustive research, the author does not consider the frequency of Dutch loanwords, merely whether some words might have become obsolete. One reason why the frequency of a loanword matters is to measure the word's vitality in a language.

This thesis will investigate the frequency of Dutch loanwords in American English and will evaluate whether frequencies have increased or decreased over time. To test this, the *Corpus of Historical American English* (*COHA*) and the *Corpus of Contemporary American English* (*COCA*) will be consulted. From these corpora I am hoping to glean findings that give insight into the distribution of words over time and the domains in which they are most productive. This thesis will try to answer the following question: Does the frequency of Dutch loanwords in American English increase or decrease over time? This

research will attempt fill the gap of a corpus-based study of the frequencies of loanwords, whilst also touching on the social meaning of loanwords. This latter notion can be linked to the bigger picture of sociolinguistics.

2 Literature Review

This literature review will give a brief introduction to the process of borrowing and the concept of loanwords in general. Additionally, adaptations of and difficulties with loanwords will be discussed, alongside the social meaning that loanwords carry. This review will then give a concise historical background of the Dutch in (what is now) northeastern North America to help better understand the origin of Dutch loanwords in American English and how they emerged. Next, this section will zoom in on Dutch loanwords in American English, discussing the different categories of American English in which Dutch loanwords occur and how the words have been preserved. Other cases of Dutch in American English will be brought up, such as calques. Lastly, previous research regarding Dutch loanwords in American English specifically will be discussed in the context of its relevance to my research.

2.1. Introduction to loanwords

Loanwords are words which are "adopted or borrowed from another language" (*oed.com*). Loanwords are also called borrowed words, which are the outcome of the process of borrowing. Treffers-Daller (2010) defines 'borrowing' as incorporating features of one language into another (p. 17). Schultz (2017) describes it as an operation through which new linguistic material is adopted from a different language (p. 44). Zenner et al. (2019) add that borrowing is the result of a language-external process, where a source language (input) and a receptor language (receiving) come into contact with each other (p. 1). Poplack, Sankoff and Miller (1988) state that the main purpose of borrowing is to define technologically or culturally new notions, or to refer to already identified concepts in a new way (pp. 47-48). The terms 'borrowed words' and 'loanwords' can be used interchangeably, but this thesis will refer to this type of words as loanwords.

Loanwords occur when multiple languages come into contact with each other. Van der Sijs (1998) explains: You get to know new things through contact with others. And those things need a name. What, then, is easier than to adopt the name the other uses? This is how

loanwords make their entrance1 (p. 9). The language contact between the Dutch and the Americans dates back to the seventeenth century. The historical background of this contact will be discussed later in section 2.2. of this literature review.

Loanwords play an essential part in the development of a lexicon by expressing new notions or expressing existing notions in new ways (Durkin, 2014, p. 428). Loanwords also have a certain status, which makes them permanent words of a language, rather than a temporary unit. This status is linked to the fact that loanwords recur regularly, are universally employed in the speech community and have gained a specific level of acceptance or recognition (Poplack, Sankoff and Miller, 1988, p. 52). While the influence of loanwords on a language's core vocabulary is hard to accurately specify, it is clear that loanwords have increased the available pool of lexical variation (Durkin, 2014, p. 421).

2.1.1. Adaptations of loanwords

Loanwords undergo multiple adaptations when they are adopted into a language. Schultz (2017) describes how adaptation "is used to refer to the naturalization which a lexical item from a foreign language underwent in order to become integrated into the linguistic system of the target language. The term also specifies the assimilated word itself" (p. 44). One specific type of adaptation a loanword undergoes is the phonological shape. Poplack, Sankoff and Miller (1988) explain that the phonological form of a borrowed word is adjusted to native patterns (p. 70). As Dutch and (American) English are both (West) Germanic languages (see section 2.1.2 below), this might predict that the Dutch loanwords are only slightly adjusted.

The concept of age also plays a role in a loanword's adaptation. Both the age of the speaker and that of the loanword. Concerning the speaker's age, multiple studies (Haugen 1950; Poplack, Sankoff and Miller 1988; and Thomason and Kaufman (1988)) have provided evidence for the fact that older speakers incorporate loanwords into the phonological shapes

of the borrowing language more than younger speakers. Regarding the age of the loanword, Poplack, Sankoff and Miller (1988) point out that the most extensive "phonological adaptation to recipient-language phonology" is found in earlier loanwords, while the recent loanwords stay more akin to their models (p. 70). This shows that while loanwords are close to the original word of the language from which they are borrowed, they are not always aimlessly copied, with no adaptation.

2.1.2. Difficulties with establishing loanwords

When it comes to loanwords, it can sometimes be problematic to ascertain a word's origin and etymology. Often, the language of origin cannot be accurately determined or words have input from multiple languages (Durkin, 2014, p. 28). Establishing the language of origin can be difficult as some loanwords may have been borrowed from one language, for example Low German, but are considered loanwords from another language, for example Dutch; and vice versa. When discussing loanwords with input from multiple languages, Durkin (2014) uses Dutch as an example. In the case of Dutch loanwords, input may come not only from Dutch as spoken in Europe, but also Dutch as spoken in South Africa and even the separate language of Afrikaans, which evolved from the Dutch vernacular (p. 28).

Another factor that may bring complications when considering loanwords is language families. These are "group[s] of languages deriving from a single ancestor or parent" (*oed.com*). Both English and Dutch belong to the same language family, namely the (West) Germanic languages, so they share a common ancestry and the two languages are relatively similar. This can complicate matters when establishing whether actual borrowing took place (Durkin, 2014, p. 49, 66). Both above-mentioned difficulties need to be taken into consideration when studying Dutch loanwords in American English.

2.1.3. Social meaning

The use of loanwords is not solely a way to use a shorter word instead of a longer one, to fill a lexical gap or a lexical act. The use of loanwords should also be viewed as a statement

to express oneself, one's social identity and one's language regard, i.e., a socially purposeful act (Zenner et al., 2019, p. 1-2). "[L]inguistic alternates within the repertoire serve to symbolize the differing social identities which members may assume", thus Blom & Gumperz (2000, p. 123). Here, these authors state that loanwords, amongst other linguistic features, bear social meaning.

This social meaning is characterised by the social characteristics and qualities related to the language feature and, essentially, its users (Walker et al., 2014, p. 169). Backus (2012) adds: "As for loanwords, in a […] sociolinguistics account, the use of foreign units […] would be seen as raising their degrees of entrenchment in the mental representations of individual speakers" (p. 6). The social meaning of loanwords can be studied from a sociolinguistic point of view. Sociolinguistics is the "branch of linguistics that deals with the social aspects of language" (*oed.com*).

The phenomenon of language variants, such as loanwords, being linked to social attributes is known as indexicality. The indexical significance of a language component regards the component's capability to index, i.e., evoke, social aspects of the context and speaker of the component. Thus, by using specific language variants, in this case loanwords, in distinct circumstances, language users disclose their language regard. Preston (2013) explains this language regard as a speaker's belief systems and cultural knowledge with regards to the social meaning of the language varieties and components in their repertoire (p. 96). According to Backus (2012), the social meaning of, for example, loanwords depends on both community-based factors and speaker-based factors (p. 8). The former factors include the dominance of one of the two languages involved and the intensity of contact between these two languages. The latter factors concern social class, gender or age (Backus, 2012, p. 8).

To assess the social meaning of loanwords, Backus (2012) suggests attitude measurements, such as focus groups or acceptability tasks (p. 19). Whilst these measurements would result in uncovering the social meaning, corpus studies will also supply valuable data (Backus, 2012, p. 19). This research will focus on corpus studies rather than attitude measurements.

The above-mentioned studies are relevant to what I am trying to find out about the social meaning of loanwords, as they help explain why loanwords are used, for example to

articulate one's identity, and how they affect the user, by, for example, establishing the user's language view.

2.2. Historical background

Studying the history of loanwords and the historical context in which borrowing took place uncovers additional aspects on the historical development of a culture and a society (Durkin, 2014, p. 428). Backus (2012) adds that outlining the historical background yields information about historical loanword layers (in this research: Dutch words in American English) and about the cultural scenarios that are inferred (in this research: speakers of American English in northeastern America adopting words from the Dutch-speaking immigrants) (p. 4).

The existence of Dutch loanwords in American English can be explained through the history of the Dutch and the Americans and how the histories of these countries and their people intertwined. During the seventeenth century, Dutch immigrants from the Low Countries (now: The Netherlands) arrived in America, specifically the northeast coast of the (now) United States, to trade. This created the first instances of language contact between Dutch and American English. Neighbouring countries have common borders, but contacts also take place over greater distances - through trade and immigration2 (Van der Sijs, 1998, p. 9). With English as a language expanding geographically, connections with other European languages happened in places outside of Europe (Durkin, 2014, p. 353). The Dutch in America is a clear example of these connections: contact still occurred despite the great distance between both countries and their people.

The main event which led to the Dutch language arriving in North America, was in September 1609. Henry Hudson, an English captain in the service of the Dutch Republic, sailed up the river Hudson on a ship called *De Halve Maen* (*The Half Moon*), which was owned by the Dutch *Vereenigde Oostindische Compagnie* (VOC) (*East India Company*)

(van der Sijs, 2010, p. 20). The exploration of Hudson allowed Dutch merchants to set up fur-trading expeditions in the regions (Jacobs, 2009, p. ix). The colonised area was called New Netherland as it had many "worthy appurtenances corresponding with the Netherlands […] being […] a newfound Netherland" (Van der Donck 1655).

After the Dutch traders had established the West Indian Company, in 1621, the colonists settled permanently, in 1624, in present-day Connecticut, Delaware, New Jersey, New York, and Pennsylvania, making New Netherland a Dutch province with Dutch as its official language (Van der Sijs, 2010, pp. 20, 27). New Netherland was part of the New World, which was the "continent […] discovered or colonized at a comparatively late period; *esp*. […] the continents of America" (*oed.com*).

Fort Amsterdam was built in 1625 on the southern part of the island of Manhattan to protect the colonists from possible attacks (Van der Sijs, 2010, p. 24). Fort Amsterdam turned into the centre of the town of New Amsterdam, which is now New York City. The era of Dutch dominance in eastern North America did not last long, as the Dutch director-general Peter Stuyvesant had to surrender the province to the English when they sailed into New Amsterdam in 1664 (Van der Sijs, 2010, p. 26). This combination of the Dutch settling in America for trade and colonising the area, resulted in many loanwords which display items of Dutch manufacture and Dutch colonial and trading activities outside Europe (Durkin, 2014, p. 357-358).

Knowing the historical background of the Dutch and the Americans is crucial in understanding how Dutch loanwords in American English emerged.

2.3. Dutch loanwords in American English

With the contact between the speakers of Dutch and American English, came the language contact between Dutch and American English. When the Dutch arrived in North America, they encountered unfamiliar objects, which they named. These objects included "specific geographical circumstances, indigenous animal and plant species, individual social and cultural developments, and technical innovations" (Van der Sijs, 2010, pp. 12-13). These areas hint at the different categories of life in American English that the Dutch influenced,

which will be discussed later in section 2.3.1. of this literature review. As the Dutch were the first Europeans to name some of these objects, the Dutch names have usually sustained until today (Van der Sijs, 2010, p. 21). Dutch loanwords stem from Dutch immigrants in the US and can therefore be attributed to immediate contact between speakers of Dutch and American English in North America (Van der Sijs, 2010, p. 116). This proves the linguistic influence of Dutch on American English. As Durkin (2014) puts it: "Many loanwords reflect contact between Dutch and English speakers outside Britain. Borrowings in North America reflect both the language of the early Dutch colonists and the Dutch spoken by subsequent waves of immigrants" (p. 358).

2.3.1. Categories

The Dutch loanwords in American English can be categorised into multiple semantic fields. Most researchers made up their own categories (Llewellyn 1936, Van der Sijs 2010) or used a combination of their own categories and those provided by corpora (Schultz 2012, 2017, 2018, 2019). Schultz has not done research specifically on the Dutch influence on American English, but the author has researched other languages, such as French, Japanese, German, Spanish and Yiddish and their influence on English. For this thesis, I have looked at the different categories these researchers used and I have renamed some and merged some which were very similar. See Appendix 3 for an overview of all the (relevant) categories used by the corpora and authors significant for this thesis. I will elaborate on why and how I did this in section 4.2. of the results section.

A couple of categories are prominent and encompass the most loanwords. Trade is one category that stands out. As explained previously, the Dutch were renowned traders. This resulted in many Dutch loanwords in American English being related to trade. There is, however, a time dimension to this: especially the earlier loanwords tend to be related to trade more often than the later ones (Llewellyn 1936, pp. 174-175). Loanwords that fall into the trade category are words for money and units of measure, and transport by sea and land (Van der Sijs 2010), but also words related to commerce and imported cloth (Llewellyn 1936).

Words associated with imported cloth can also be linked to the category of fashion and lifestyle (Llewellyn 1936, Van der Sijs 2010, Schultz 2012). The loanwords regarding trade go both ways: American concepts which were novel to the Dutch and vice versa:

> Dutch immigrants […] trade[d] with the […] Americans. They brought furs back with them to Europe, and in exchange they took all kinds of Dutch products with them to the New World. Much of that was new to the […] Americans, who adopted the Dutch terms for these items, adding them in their languages. (Van der Sijs, 2010, p.12)

Language and linguistics is another category that encompasses many Dutch loanwords (Llewellyn 1936, Van der Sijs 2010, Schultz 2012). A specific language group is children's language, which includes words related to children's games such as *hunk* (meaning 'home base') and *scup* (meaning 'swing'). According to Llewellyn (1936), it is sensible that children's language is a prominent category, as English and Dutch children played together, thus influencing each other's languages (p. 178).

The church played an important role of the preservation of Dutch in eastern North America (which will be discussed in more detail later in section 2.3.2. of this literature review), so it is no surprise that faith and religion is another prominent category (Van der Sijs 2010, Schultz 2012).

As explained earlier, when the Dutch arrived in North America, there were many new objects which they named. These objects are words which appear in categories such as flora and fauna (including fishing). This is because when the Dutch arrived in Northeastern United States, they were confronted with a natural environment for which they did not yet have words (Van der Sijs, 2010, p. 144).

Another category with a large number of loanwords is gastronomy (Van der Sijs 2010, Schultz 2012, 2017, 2018). Van der Sijs (2010) argues that this category is where the most considerable contribution of Dutch to American English took place (p. 117). Especially the terms for sweets and confectionery were borrowed quite extensively. This category includes words related to food, drink and stimulants. Van der Sijs (1998) specifies drink-related words: Dutch beer, brandy and jenever: export products par excellence3 (p. 127).

This category is related to that of household effects and everyday implements (Llewellyn 1936, Van der Sijs 2010), which includes objects in and around the house, and everyday household objects, especially related to cooking.

Lastly, Dutch loanwords in American English can be categorised into civilisation and politics (Van der Sijs 2010, Schultz 2012, 2017, 2018). These words regard polity and citizens. This category can be linked to military and nautical words (Llewellyn 1936), considering that politics influences the military and marine.

Categorising loanwords and studying categories used by other authors is relevant in this research to find out which spheres of life in American English were influenced by Dutch.

2.3.2. Preservation

Even though the Dutch were in control in northeastern North America for only about 40 years, the influence of Dutch on American English lasted longer and was well preserved. The main reason of the preservation of Dutch was the church. Llewellyn (1936) explains that the northeastern United States upheld much of its Dutch institutions and language for a long time, and even though most of the schools had got rid of Dutch shortly after the English conquest, the service of the Dutch Reformed Church remained Dutch until 1764 (p. 174). So people who had changed their language to English would still need to be able to understand Dutch in order to understand the church services. Dosker (1880) adds that the religious disposition of the Dutch, coupled with their conservative character, has contributed much to the preservation of the Dutch language in the East4 (p. 46).

The fact that Dutch was preserved relatively long and well helps with this research, as it ensures that Dutch loanwords are somewhat embedded in American English.

2.3.3. 'Dutch' loanwords in American English

Some words, which are claimed to be Dutch loanwords, do not actually occur in Dutch dictionaries. Van der Sijs (1998) gives some examples: *ianthinite* 'kind of uranium oxide', derived from Dutch *janthiniet*, formed by A. Schoep in 1926 from Greek *ianthinos* 'violet coloured'; *javanite* 'Javanese tektite' from Dutch *javaniet*, formed in 1936 by Von Koenigswald; *stainierite* 'sort of cobalt oxide' from *stainierit*, formed by V. Cuvelier in 1929 from the name of the Belgian geologist Xavier Stainier; and *truscottite* 'kind of silicate' from *truscottiet*, formed in 1914 by P. Hövig from the name of the English mining engineer S.J. Truscott5 (p. 153). These words may be part of American English, and make use of Dutch word-formation strategies, but they are not part of Dutch, therefore are not Dutch loanwords.

2.3.4. Dutch loanwords in 'American English'

Similarly, some words occur in Dutch dictionaries, but were not actually fully integrated into American English. Several lists of Dutch loanwords have circulated and have been carelessly copied. These words include "*connalyer* “crowd” (from Dutch *canaille* “mob”), *coss* “chest (of drawers)” (from Dutch *kast*), and *klainzaric* “untidy” (from Dutch *kleinzerig* “over-sensitive”)" (Van der Sijs, 2010, p. 14). These words, however, do not occur in American English, so they could be considered Dutch loanwords, but not in American English.

2.3.5. Dutch beyond loanwords in American English

Some people argue that the Dutch influence on American English did not stop at loanwords. There are other language characteristics or specific meanings in American English which can be traced back to Dutch. Jan te Winkel (1896), a Dutch philologist, argues the following:

> Furthermore, many English words occur with a meaning different from the one they have in proper English, but which is immediately understood by a Dutchman. An American will say for “I think,” *I believe*, as we Dutchmen do. When he is *sick*, he does not mean, as an Englishman does, ‘sick,” but “ill,” as we do in Dutch. “To (carry on) trade” he calls to *handle*, preferring *hard money* (hard geld, we Dutchmen say) to paper money, although he will accept both *all two* (Dutch *alle twee*); but before a purchase he likes to indicate that actually *the whole boodle is worth not a red cent* (Dutch *de hele boel is geen rooie duit waard*); for if he looked too eager beforehand he would be a *muttonhead* (Dutch *schaapskop*). If he has visitors, *folks* (Dutch *volk*), it is sometimes *allmighty full* (Dutch *allemachtig vol*) in his house. (pp. 343-344)

These examples are called calques, which are "loan-translations […] in more or less literally translated form" (*oed.com*). Calques arise during a process in which word 'x' in language A that translates the primary meaning of word 'y' in language B copies a secondary meaning of word 'y' (Geeraerts, 2009, p. 36). Whilst this could reveal that Dutch did not only influence American English through loanwords, but also through other linguistic aspects, it is a source from 1896, therefore not very reliable anymore. Additionally, whilst some of these examples still occur today, such as *folks*, not all of these phrases are current today or appear anymore.

2.4. Previous research

2.4.1. Schultz

Schultz (2012, 2017, 2018, 2019) has investigated the influence of numerous languages on the English language using a consistent framework. This framework leads to the discovery of words and meanings with a foreign language’s etymon in their etymological descriptions. Schultz’s (2012, 2017, 2018, 2019) framework also presents the date of entry of words. The framework provides this option so the user can identify which words entered a language when. This is the author’s focus, rather than trying to contribute to more theoretical considerations.

Schultz's (2012, 2017, 2018, 2019) studies are relevant to what I am trying to find out as they provide a useful framework to find out which words in (American) English contain a Dutch etymon. Furthermore, using this framework will yield insight on the influence of Dutch on the different lexical domains of American English. Employing Schultz's categorisation will illustrate different fields of life and subject areas in American English to which Dutch has contributed new meanings and words.

2.4.2. Van der Sijs

Van der Sijs (2010) has analysed the influence of Dutch on the North American languages. The author glosses over 300 words in *Cookies, Coleslaw, and Stoops* (2010). This glossary includes the loanword, its meaning and the first recorded occurrence. Van der Sijs (2010) not only examines Dutch loanwords in American English, but also in Native American languages. This thesis will not focus on these languages, only on American English, but with this research, Van der Sijs (2010) proves that Dutch has influenced multiple languages through loanwords.

Applying Van der Sijs' (2010) framework is relevant to this research as it provides an exhaustive list of Dutch loanwords in American English. Furthermore, Van der Sijs' (2010) glossary gives insight into the definitions and origins of the loanwords, the context in which the words are used and time periods relevant to the loanwords. This information is significant when analysing the frequency of the loanwords. It needs to be clear what the meanings of the loanwords are, so they are not confused with other meanings of the same word. Additionally, knowing a loanword's first recorded occurrence helps to know if that year is prior to or during the years provided by *COHA* and *COCA*.

3 Methodology

To investigate whether the frequency of Dutch loanwords in American English increases or decreases over time, Schultz's framework (2012, 2017, 2018, 2019), using the *OED*, was utilised. The outcome of loanwords of this framework was compared to van der Sijs' (2010) glossary of loanwords. To allow for comparison, solely loanwords that appeared in both lists were used for this research. To further trace the development of these loanwords over time, two corpora were employed to consider the frequency, namely *COHA* and *COCA*. This methodology section will give an overview of what the frameworks and corpora used for this research entail and how these were applied for this research.

3.1. Schultz's framework

Schultz (2012, 2017, 2018, 2019) has researched the impact of multiple languages on the English language. Whilst the languages and their impact on the English language Schultz researched differed, the author's framework remained the same for each study. Schultz used the *OED Online* to perform her research. This online version of the dictionary provides an advanced search option which allows the user to find all the words and meanings with a foreign language's etymon in their etymological descriptions. The advanced search option also allows the user to enter a specific year or time period in the date of entry of the words. Schultz (2012, 2017, 2018) focuses on the twentieth century, so from 1901 onwards. Schultz (2012, 2017, 2018, 2019) divides the results into categories, which are a combination of the classification of the *OED* and her own categorisation. Examples of categories Schultz (2012, 2017, 2018, 2019) uses are gastronomy, fashion and lifestyle, and technology. See Appendix 3 for a complete list of Schultz's (2012, 2017, 2018, 2019) categories.

I will employ Schultz's framework by combining it with Van der Sijs' framework. What the combination of these two approaches looks like and which elements I use from the respective frameworks is explained in section 3.4. in this methodology section.

3.2. Van der Sijs' framework

Van der Sijs (2010) has researched the influence of Dutch on the North American languages. The glossary of Dutch loanwords in American English she provides in *Cookies, Coleslaw, and Stoops* (2010) is based on words which have been derived from American English dictionaries. The glossary includes words that are still in use, even though some have turned into historical terms. The loanwords are followed by the contemporary meaning and concise information about the origin of the American English word. Additionally, the time span in which the loanword was presumably borrowed from Dutch and details on how widespread or common the word is, are given. Furthermore, the first recorded occurrence of the word in a quote is given, if possible, with a source. Van der Sijs (2010) argues that the Dutch loanwords can be divided into thirteen categories, or semantic fields as the author calls them (2010, p. 116). Examples of categories Van der Sijs (2010) uses include food, drinks, and stimulants; household effects and everyday implements; and human traits and characterisations. See Appendix 3 for a complete list of Van der Sijs' (2010) categories.

Like Schultz (2012, 2017, 2018, 2019), Van der Sijs (2010) also uses categories. I will use a combination of both authors' categorisation, in addition to Llewellyn's (1936) categories and categories offered by the *OED*. How and why I used this combination will be explained in section 4.2. in the results section.

3.3. Corpora

This thesis will use three main corpora to conduct the research and answer the research question. These corpora are the *Oxford English Dictionary* (*OED*), the *Corpus of Historical American English* (*COHA*) and the *Corpus of Contemporary American English* (*COCA*).

3.3.1. The *Oxford English Dictionary* (*OED*)

The *OED* is the dominant historical dictionary of English, published by Oxford University Press. The first fascicle was published in 1884. In 2000, the online version of the dictionary

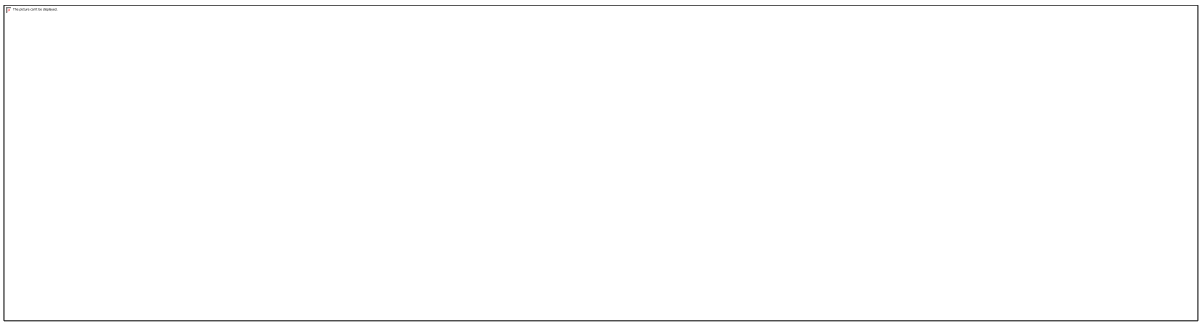
was made available. The *OED Online* is updated, revised and extended at regular intervals. For this thesis, solely the online version will be used, so when referring to the *OED* it will always mean the *OED Online*. The *OED* is searchable online at http://www.oed.com.

For this thesis, I used Schultz's (2012, 2017, 2018, 2019) framework using the *OED*, with some minor adjustments. The *OED* offers an advanced search function, through which the entries can be searched with refined criteria, such as the language of etymology (see Image 1) and a time period in which the entry occurred (see Image 2). This is necessary in order to find entries with a Dutch etymology from a specific period.

3.3.2. The *Corpus of Historical American English* (*COHA*)

The *COHA* is the largest structured online corpus of historical English. It contains more than 475 million words of different genres from 1820 until 2000. These genres include fiction, magazines and newspapers, reflecting a range of formal and informal text types. *COHA* is searchable online at www.english-corpora.org/coha/. For this research, *COHA* was used to gain an insight into the frequency of Dutch loanwords in American English in an earlier time period, i.e., during the nineteenth and twentieth century.

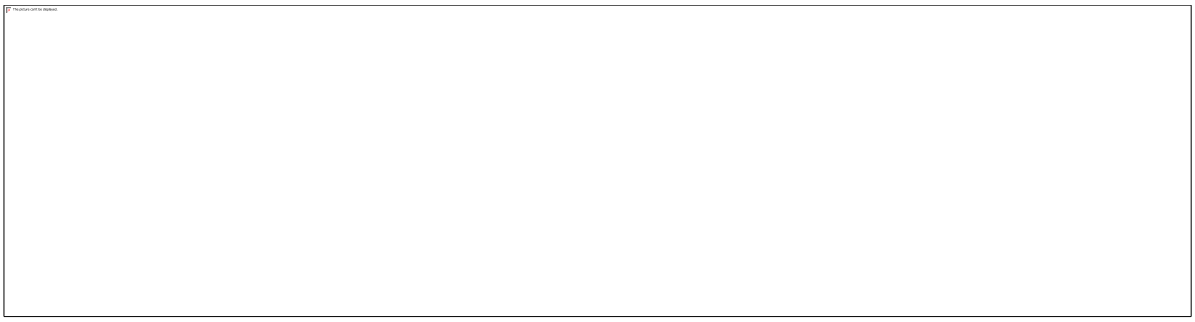
See Image 3 for a representation of the bar charts in *COHA*.



3.3.3. The *Corpus of Contemporary American English* (*COCA*)

The *COCA* is an online corpus of contemporary American English. It consists of one billion words and almost half a million texts. *COCA* contains texts from 1990 until 2019. These texts are both formal, such as newspapers and academic texts, and informal, such as movie subtitles and blogs. *COCA* is searchable online at www.english-corpora.org/coca/. For this research, *COCA* was used to give an indication of the frequency of Dutch loanwords in American English in a later time period, i.e., during the late twentieth and early twenty-first century. I will also be taking into account the role of the text-types: both between text-types (formal versus informal) and within text-types (e.g., newspapers and academic journals (both formal) or blogs and webpages (both informal)).

See Image 4 for a representation of the bar charts in *COCA*. The charts to the left of the dark blue bar are the different text-types. The light grey charts represent informal text-types: blogs (BLOG), webpages (WEB), and tv shows and movies' subtitles (TV/M). The light blue charts to the left of the dark blue bar illustrate formal text-types: spoken (SPOK), fiction (FIC), popular magazines (MAG), newspapers (NEWS), and academic journals (ACAD). The charts to the right of the dark blue bar are the years, divided into sections of four years. *COCA* does not give the total frequency per million words for all years, but for periods of four years. This corpus does provide the total frequency per million words for all text-types and per text-type. I have calculated the total frequency per million words for all years myself, by adding the frequency given for each period of four years, and dividing that by six (as there are six columns) (see results in Appendix 2).



3.4. Methodology

For this research, the above-mentioned frameworks and corpora were combined and applied. Firstly, the advanced search from the *OED*, used by Schultz (2012, 2017, 2018, 2019) was filled in as follows: Entries containing *'Dutch'* in *'Etymology'* and *'1820-'* in *'Date of entry'*. This would result in English words that were borrowed from Dutch from 1820 onwards. 1820 was chosen as the start year, as that is the year the texts in *COHA* start. The list of loanwords that resulted from the *OED* was compared to Van der Sijs' (2010) glossary of Dutch loanwords in English. To provide room for contrast, the words that occurred in both lists were used exclusively.

The year the *OED* gives, next to the entry, is the year of the first recorded quote in which the word appeared. Van der Sijs' (2010) glossary gives the first occurrence of each word in a quote, which was used to compare to the year the *OED* provided (see Appendix 1). My reason for comparing the year the *OED* supplies with the year in Van der Sijs' (2010) glossary is to attempt to determine that it concerns the same meaning of a word (see section 4.2.2. for an example). The time period in which loanwords were borrowed can give more information about the situation in which a word is borrowed. For example, loanwords from the seventeenth or eighteenth century are undoubtedly derived from the first Dutch settlers. Van der Sijs (2010) argues that only relative value should be assigned to the loanwords' first occurrence, as printed texts and dictionaries are usually late with adopting specific words (p. 115). However, the first recorded instances are also highly informational. The oldest occurrences illustrate a word's original usage, whilst more recent instances demonstrate how the word is used in the current language. This highlights the relevance of my corpus study. This corpus study investigating frequencies yields additional information

regarding loanwords' usages, whether they are well-embedded in a language and whether they are still used today.

The words that occurred in both the *OED* and Van der Sijs (2010) were firstly entered in *COHA* and then in *COCA*. The words were put in the search bar of both corpora to result in frequency lists. The frequency lists not only indicate a word's frequency, but also present the distribution of the word frequency across different decades and different text-types (such as formal versus informal). The Dutch loanwords were entered into both *COHA* and *COCA* and, with the use of the given bar charts, the course of the frequency was analysed.

As mentioned in section 4.2.1. in the results section, *COHA* and *COCA* were sensitive to one spelling variant, namely using a hyphen or not. This had slight implications for my method, as it meant I had to enter the affected loanwords (*bed*(-)*spread* and *beer*(-)*hall*) twice: once with each spelling variant. I made use of one parametre the corpora allow, namely the parts of speech (POS) search option (see section 4.3.1.3. in the results section for elaboration). This parametre enables the user to search for categories of lexical items with similar grammatical properties.

This survey will tell me the course of the frequency of the selected Dutch loanwords in American English. From this course, I am hoping to analyse the recurrence and draw potential conclusions regarding, amongst others, the extent and life-span of Dutch loanwords in American English.

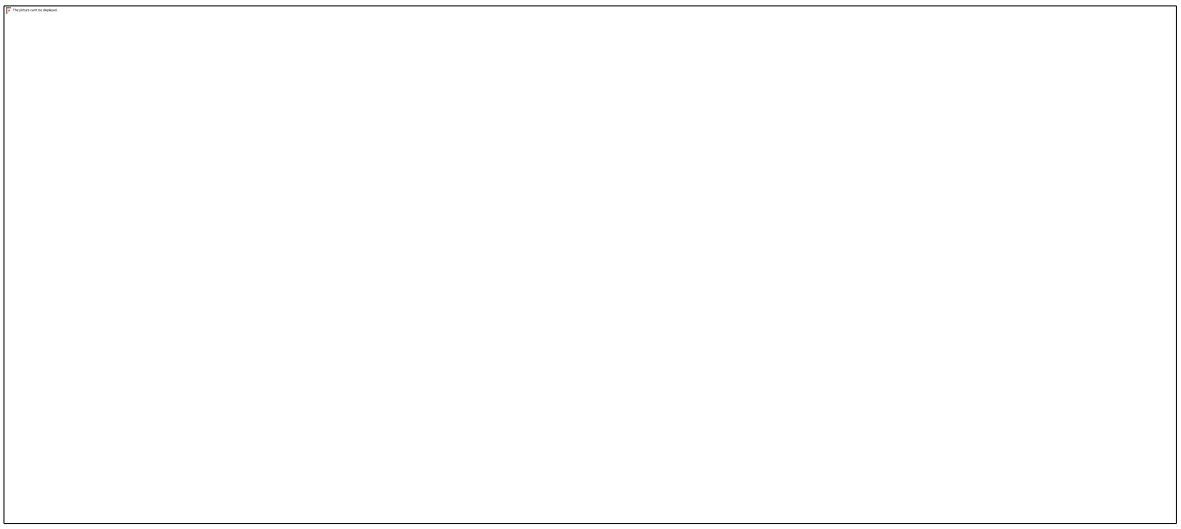
4 Results

The results discussed below contribute to my research aim of analysing the frequency of Dutch loanwords in American English and establishing whether the frequency increases or decreases, using different frameworks and corpora as my research approach. I have included a set of tables and screenshots to summarise some of my findings and illustrate some of the results I discuss. A complete overview of all results can be found in the appendices. The results are presented in normalised frequencies (per million words) as well as raw frequencies (in brackets). The former is used to get an accurate understanding of the frequency, as the corpora differ in size, so this gives the normalised frequency. The frequency per million can sometimes be 0.00, even though a loanword still resulted in one hit. This is due to the rounding off of the digits after the comma for the frequency per million words. I have selected a couple of loanwords, which stood out for various reasons. I will mention those loanwords in this section, which will be discussed and analysed in further detail in the discussion section. I will start with the results from the *OED* and Van der Sijs (2010), then I will elaborate on the categories I have put the loanwords into and present the data that resulted from my *COHA* and *COCA* survey.

4.1. *OED* and Van der Sijs (2010) results

The corpus data on which these results are based were retrieved from the *OED* during the months of March, April and May in 2022. During this time, the search results did not include any Dutch loanword which had been borrowed from Dutch in the twenty-first century. The Dutch loanwords collected through the *OED*'s advanced search amounted to 643 lexical items. Van der Sijs' glossary (2010) contains 246 lexical items. The majority of these lexical items are still in use, but to different degrees. The minority of the loanwords are only used as historical terms, such as *barraclade*; some words are solely used regionally, for example *dobber*; and some are widely used, for instance *bed(-)spread* (Van der Sijs, 2010, p. 13). Thirty-seven Dutch loanwords occurred both in the *OED* when applying Schultz's (2012, 2017, 2018, 2019) framework and in Van der Sijs' (2010)

glossary. See Table 1 for an overview of these 37 loanwords that were used for this research.



4.1.1. Spelling variants

Some words occurred in both the *OED* and Van der Sijs' glossary (2010), but were spelled differently. One spelling variant is the use of a hyphen: Van der Sijs (2010) uses a hyphen, where the *OED* does not. This was the case for two loanwords: *bed-spread* (Van der Sijs 2010) versus *bedspread* (*OED*) and *beer-hall* (Van der Sijs 2010) versus *beer hall* (*OED*). Interestingly, the only quote Van der Sijs (2010) provides for *beer-hall* follows the spelling without hyphen: "Bicyclists who are making the concert gardens and *beer halls* in the suburbs flourish" (p. 234). The *OED* presents three variants of spelling for *beer hall*: "Then there were the great *beer-halls*" (*oed.com*), "[t]he *beer hall* still hosts Burns Night celebrations" (*oed.com*) and "chanting old Bavarian toasts in a crowded *beerhall*" (*oed.com*). The quotes for *bed-spread* provided by Van der Sijs (2010) all follow the spelling with hyphen, for example: "I made a *bed-spread* of his skin, and the way it used to cover my bar mattress … would have delighted you" (p. 171). The *OED* provides quotes with both spellings: "I made a *bedspread* of his skin" (*oed.com*), but also "*Bed-spread* […] the common name for a […] coverlet" (*oed.com*).

Another spelling variant between both lists is capitalisation: where Van der Sijs uses a capital letter, the *OED* does not. This was the case for two loanwords: *Kabouter* (Van der Sijs 2010) versus *kabouter* (*OED*) and *Provo* (Van der Sijs 2010) versus *provo* (*OED)*.

Whilst Van der Sijs (2010) capitalises the first letter of *Kabouter*, the given quote maintains the other spelling: "[T]he *kabouters* were best known […] for their […] scheme" (p. 279). The *OED* provides quotes with both spellings: "Their successors, the *Kabouters* […] have gone one better" (*oed.com*) and "[w]hat city would have a political movement called the *kabouters*" (*oed.com*). Regarding *Provo*, Van der Sijs (2010) does not give a quote. The *OED* distinguishes between the name of the movement and another meaning. For the former definition, the word is capitalised: "*Provo*, the Dutch movement" (*oed.com*). For any other meaning, a lower-case *p* is used, as in "you don't look like […] a *provo*" (*oed.com*).

These spelling variants may occur as it regards 'older' loanwords and copying these words over time may have resulted in different variations. Regarding capitalisation, it may have been hard to establish whether a lowercase or uppercase letter was used, especially in the cases of 'k' and 'p', where there is no indisputable difference. Concerning hyphens, they might have been left out for the convenience. See the implications spelling variants had for my method in section 3.4. of the methodology section.

4.1.2. Year differences

Not only do the *OED* and Van der Sijs (2010) differ in spelling, but also in the year of the oldest known quote. For most loanwords, such as *barraclade* (1848), both the *OED* and Van der Sijs (2010) give the same year of the oldest known quote. The majority of the loanwords have a difference of 10 years or less, such as *dingus* (1866 (*OED*) versus 1876 (Van der Sijs)) and *adstratum* (1939 (*OED*) versus 1932 (Van der Sijs)). However, for some words there is a difference of more than 50 years, such as *advocaat* (1895 (*OED*) versus 1945 (Van der Sijs)) and *banket* (1886 (*OED*) versus 1982 (Van der Sijs)). This proves that what is known about the year of the oldest known quotes of words is not always reliable.

It might also be the case that different meanings of the same word have different years of first recorded occurrence. For example, the *OED* gives two definitions for *bazoo*: 'trumpet' (from Dutch *bazuin*) and 'mouth'. The first definition has a first recorded occurrence in

1877, as in: "Blowin' his bazoo" (*oed.com*). Typically, this is called different sense of the word, but in this case, it concerns completely unrelated meanings. The second definition, on the other hand, first reportedly occurred in 1948, in the sense of: "Shut yer big bazoo!" (*oed.com*). For this research, I have made sure to check that the *OED*'s and Van der Sijs' (2010) definitions corresponded. See Appendix 1 for an overview of all the years of the first recorded quotes containing the Dutch loanwords.

4.2. Categories

For this research, I divided the loanwords into different categories, which were based on a combination of the categories provided by the *OED*, the categories Llewellyn (1936) uses, Van der Sijs' categories (2010) and Schultz's categorisations (2012, 2017, 2018, 2019). See Appendix 3 for all the categories used by the *OED* and these authors. Some categories were merged, whilst others were renamed. I renamed and merged the categories, as some were very similar, but had a different name. For example, 'consumables' (*OED*); 'food, drink, and stimulants' (Van der Sijs 2010); and 'gastronomy' (Schultz 2012, 2017, 2018, 2019). By renaming and merging multiple categories, I narrowed the 68 different categories down to 11 overarching, relevant ones. Unfortunately, the *OED* did not provide a category for each loanword; in those cases, I based my category on Van der Sijs' (2010) one and related categories from Llewellyn (1936) and Schultz (2012, 2017, 2018, 2019).

Based on the definitions and explanations given by the *OED* and Van der Sijs (2010), the Dutch loanwords were divided into 11 categories. These categories can be referred to as the spheres and fields of life of American English that were influenced by Dutch. The list below indicates an overview of the different categories, in descending order, with an illustrative example per category. The number in brackets signifies the amount of loanwords that fit into the concerned category.

1. Household objects (nine loanwords), e.g. *barraclade*
2. Language and linguistics (six loanwords), e.g. *dingus*
3. Food and drinks (six loanwords), e.g. *advocaat*

4. Sports and leisure (three loanwords), e.g. *korfball*

5. Human characteristics (three loanwords), e.g. *dumbhead*

6. Politics (three loanwords), e.g. *patroonship*

7. Science (three loanwords), e.g. *Kuiper belt*

8. Transport (two loanwords), e.g. *Fokker*

9. Arts (one loanword), e.g. *bazoo*

10. Flora and fauna (one loanword), e.g. *spearing*

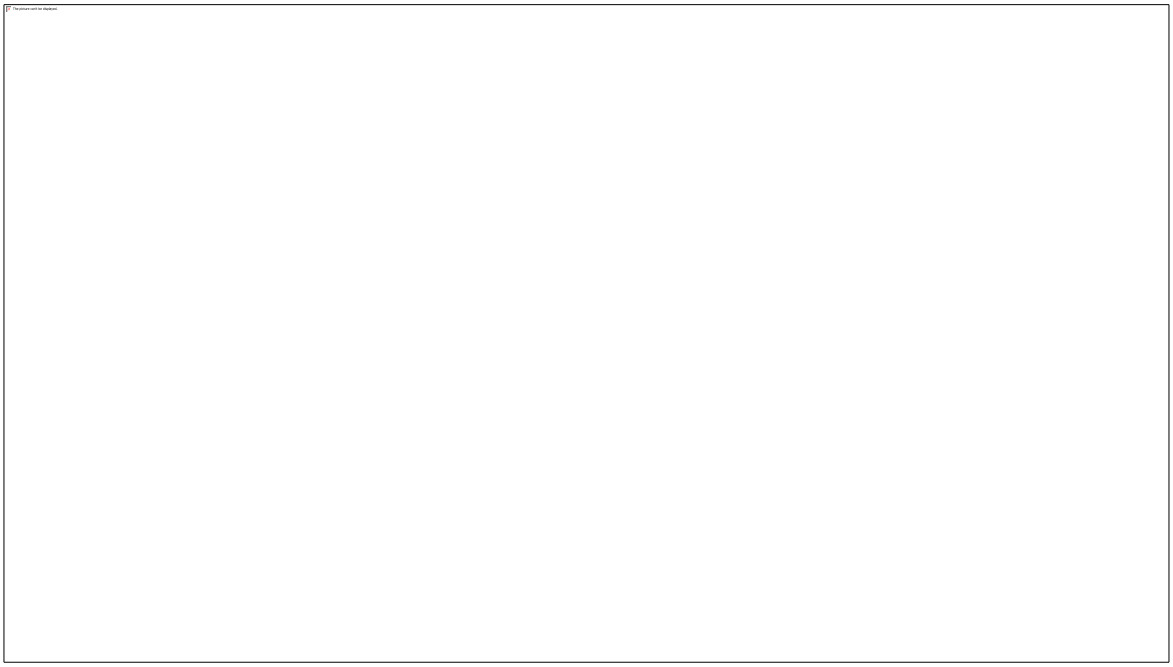
11. Miscellaneous6 (three loanwords), e.g. *poppycock*

4.2.1. Remarkable loanwords and categories

Whilst most loanwords evidently fit into one category, some words either corresponded to multiple categories or were difficult to place into one category. Loanwords that belong to multiple categories include *dobber* and *fyke*, which coincidentally were put into the same two categories: 'household objects' and 'sport and leisure'. Another loanword to which two categories were assigned is *snoop*: 'food and drinks' and 'language and linguistics'. The loanwords in the 'miscellaneous' category are *beer*(-)*hall*, *poppycock* and *woonerf*. These words did not fit into any of the above-mentioned categories, or other categories from the authors. All three authors also ran into this problem, as they have a similar category for words that do not fall into any category: 'unclassified borrowings from Low Dutch' (Llewellyn 1936) and 'miscellaneous' (Van der Sijs 2010, Schultz 2012).

4.3. *COHA* and *COCA* results

The total number of loanwords that were entered into *COHA* and *COCA* amounted to 39 words. This is two more than the initial 37 loanwords that resulted from the *OED* and Van der Sijs' (2010) glossary. These two extra words are the spelling variants regarding a hyphen. *Bed(-)spread* and *beer(-)hall* were both entered with and without a hyphen, as these different spellings resulted in distinct frequencies. Of these words, 28 words had a frequency of at least 0.01 (1) in *COHA*, versus 30 loanwords in *COCA*, see Table 2. The cells that are highlighted yellow in Table 2 are the loanwords with a frequency of at least 0.01 (1) in both corpora. Out of the 39 words, 12 were instances of a decreased frequency from *COHA* to *COCA*. Twenty-two words had an increased frequency and for the remaining five words the frequency stayed the same. For a more detailed and complete overview, see Appendix 2.



4.3.1. Differences

Whilst both *COHA* and *COCA* are related corpora with a similar set-up, they have some differences. The bar charts differ, the corpora differentiate in spelling, and the frequency outcomes also vary.

4.3.1.1. Difference in bar charts

How the bar charts are displayed differs per corpus. This is important to note, as it influences the results. The bar charts in *COHA* give sections that display the years in decades. Additionally, raw frequencies and normalised frequencies are given. The bar charts in *COCA*, however display both the frequencies across text-types and the distribution per year. See Images 3 and 4 in sections 3.3.2. and 3.3.3. of the methodology section for screenshots of the bar charts from *COHA* and *COCA*.

4.3.1.2. Spelling variants

According to the results, *COHA* and *COCA* do not differentiate between capital and lowercase letter, because the frequency remains the same. See, for example, *Provo* and *provo* (both 0.44 (178) (*COHA*) and 0.62 (457) (*COCA*) hits). However, the frequency of words with and without a hyphen does give different results: *beer-hall* and *beer hall* (0.04 (15) (*COHA*) and 0.01 (7) (*COCA*) hits with hyphen versus 0.18 (72) (*COHA*) and 0.12 (89) (*COCA*) without hyphen). See section 4.1.1. in this section for an elaborate discussion.

4.3.1.3. Difference in frequency

The frequency of the hits that resulted from *COHA* and *COCA* greatly varies per loanword. Whilst some loanwords get no hits in either of the corpora (such as *adstratum* and *rolliche*), some words get over 1,000 hits in both corpora (such as *file*: 4.27 (1,728) (*COHA*) and 8.93 (6,612) (*COCA*)). However, most loanwords with an exceptionally high frequency have multiple meanings. All occurrences in *COHA* and *COCA* are shown in their context, so a quick scan can illustrate in which situation a word is used. Browsing all hits and their context would be too consuming, regarding the loanwords with over 1,000 hits, so I looked at the first 50 to get an impression of the typical context within which the word occurs. Often the context in which the loanword has results in this high frequency was not the Dutch loanword. In the case of *file*, the 'Dutch' meaning of the word is a noun (it is a "cloth used for wiping floors" (Van der Sijs, 2010, p. 176)). To filter out *file* as a verb, the parts of speech (POS) search option was used to search for nouns only. See 5.1.1. in the discussion section for a more elaborate discussion on this issue.

4.3.2. Frequencies that stayed the same

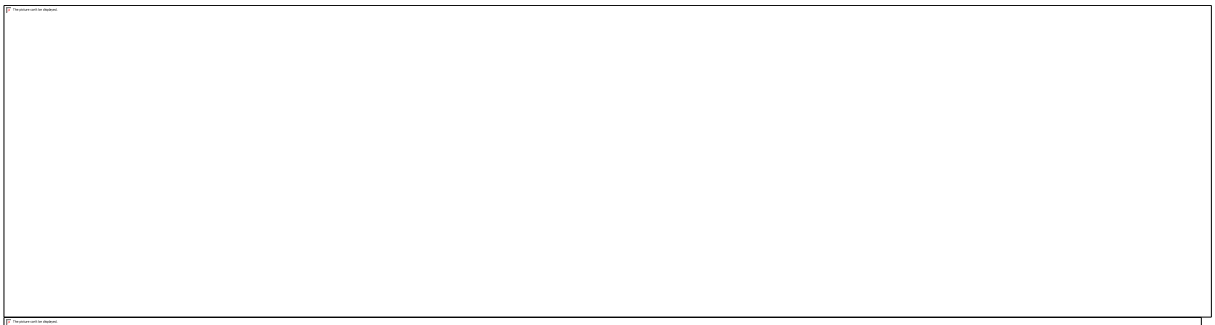
For some loanwords, the amount of hits remained the same for both corpora. Some loanwords had a frequency of 0.00 (0) for both corpora, such as *adstratum* and *barraclade*. Whilst this is a constant frequency, it could be marked as a null frequency and is therefore not very meaningful. More noteworthy, however, are loanwords that had at least one hit and remained the same, such as *dobber* 0.00 (1). In *COHA*, *dobber* is found in 1842 in fiction (*Sporting Scenes and Sundry Sketches*): "[A]nd let the dobber fall upon the water" (Herbert 1842). In *COCA*, this loanword is found in 2012 in a blog (*The Bayfield Bunch*): "Thus far I have only had dirt dobber problems in Az" (*thebayfieldbunch.com*). So whilst the frequency of the loanword stayed the same, the precise year of occurrence, context of the loanword and text-type in which it was found all differed.

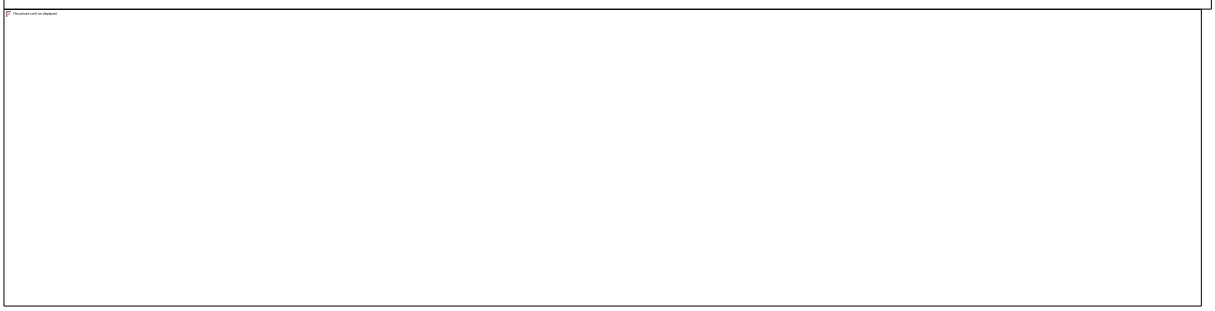
4.3.3. Development over time

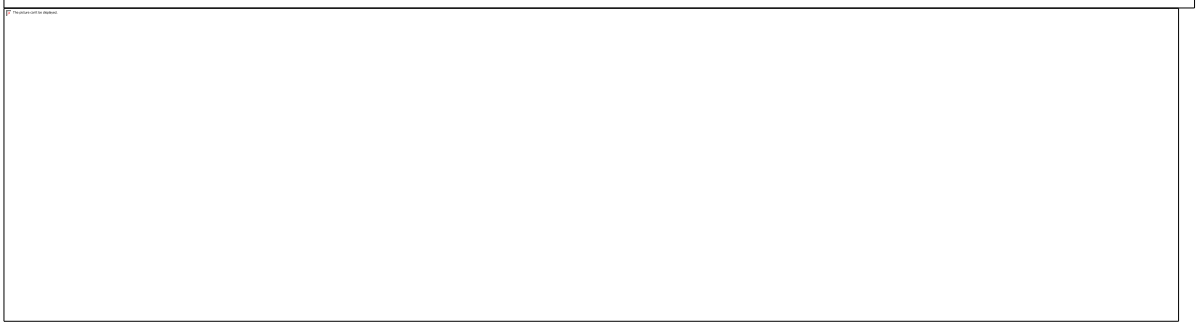
When entering a word in the search bar of *COHA* or *COCA*, a list of the total frequencies and their contexts comes up. Additionally, both corpora also offer a bar chart function. For these results, exclusively loanwords which had at least one hit in either of the corpora were used. The charts offer to look at smaller time spans, instead of all the years that the corpora cover, to consider frequency differences and state something about the development over time.

The bar charts can also be regarded as graphs to analyse the frequency over time. When analysing the chart, it is important to note that the time frames in which the bars occur are different for *COHA* and *COCA*: the former corpus presents the data in bars of 10 years, whilst the latter displays the statistics in strips of four years. Additionally, there is an overlap between the two corpora: *COHA*'s charts end in 2010, but *COCA*'s statistics already start in 1990. In order to interpret the course of the charts, these aspects were taken into consideration.

Some loanwords increase in frequency over time, which is illustrated by an ascending graph, such as *file* (see Image 5) and *Kuiper Belt*. The opposite of an ascending graph would be a descending graph in which the loanwords decrease in frequency gradually. This was not the case for any of the loanwords. A third course represented by the graphs is a parabola graph, i.e., the frequency increases, then decreases, such as *bedspread* (see Image 6) and *hunk*. Another development is a periodic graph, i.e., the frequency increases, then decreases, then increases again (or vice versa). This is the case for, for example *beer hall* (see Image 7) and *dingus*. Lastly, the frequency of some (arguably most) loanwords is a 'random' progress: there is no pattern in the frequency, but rather an irregular flow, such as *banket* (see Image 8) and *bazoo*. See images below for illustration.







4.3.4. Distributions across text-types

The distribution of frequencies across text-types yields differences. These occur between formal and informal text-types, but also within these two categories, such as tv shows and movies' subtitles versus blogs. As can be seen in Appendix 5 (screenshots from the *COCA* charts), *COCA* gives both the results in text-types and years. Whilst *COHA* does not provide the results in text-types automatically, they can be retrieved through the 'sections' option, where the user can select one or multiple text-type (both formal and informal). This research will focus on the distribution across text-types given by *COCA*, as this concerns more recent data, which makes it more accessible to research evidence that leads to conclusions.

4.4. Appendices

Appendix 1 includes the Dutch loanwords, the year in which a given word was first recorded, both according to the *OED* and Van der Sijs (2010); the difference in years given by the two sources (the *OED* and Van der Sijs (2010)); information about the word given by the *OED* and Van der Sijs (2010); the category given by the *OED* and Van der Sijs (2010); and the additional categories that I have devised for the purpose of this thesis. Appendix 2 shows the *COCA* and *COHA* frequency lists of the relevant loanwords. This appendix includes the raw frequencies and the normalised frequencies per million words in both corpora; and whether there was an increase or a decrease in hits, or whether the amount of hits stayed the same. Appendix 3 presents lists of all the semantic

categories used by the *OED* and various authors (Llewellyn 1936; Van der Sijs 2010; Schultz 2012, 2017, 2018, 2019). Appendices 4 and 5 contain screenshots from the bar charts of *COHA* and *COCA*. See section 4.3.1.1. in this section for an explanation of the data the bar charts provide.

5 Discussion

The results indicate that the majority of Dutch loanwords in American English increase in frequency over time, as more than half of the loanwords have a higher frequency in *COCA* than in *COHA*. In this discussion section, I will discuss the frequencies of the loanwords and what they represent. I will also discuss the development of loanwords over time and across text-type. I will then consider the emergence and disappearance of loanwords, and the social meaning around loanwords. Lastly, I will propose some possible future research.

5.1. Frequency

The variation in frequency was great. The lowest raw frequency was 0, which corresponds to a normalised frequency of 0.00. The highest raw frequency was 6,612, which corresponds to a normalised frequency of 8.93. Most loanwords had a frequency between 0.00 (1) and 0.50 (100). Whilst the loanwords with a frequency between 0.00 (1) and 0.50 (100) are the loanwords I will focus on, some loanwords with exceptionally low or high frequencies also deserve mentioning and I will attempt to explain their exceptional frequencies. On the one hand, frequency lists are effective, as they demonstrate whether a word is commonly used or not. The issue with frequency lists, on the other hand, is that they cannot say which meaning of a word is more commonly used, when dealing with words that have distinct definitions in varied contexts. Luckily, both corpora give the possibility to view a word in its context.

5.1.1. Exceptionally high frequencies

Loanwords that resulted in extraordinarily high frequencies, such as *file*: 4.27 (1,728) (*COHA*) and 8.93 (6,612) (*COCA*), may be explained by homonyms, i.e., words with identical spellings, but distinct (unrelated) meanings (Wales, 2014, p. 201). The Dutch loanword *file* comes from the dialectal Dutch word *feil* and is a “cloth used for wiping floors” (Van der Sijs, 2010, p. 176). However, when looking at the context in which the word is found in, for instance, *COHA*, it often has a different meaning. For example, in

COHA, *file* appeared in the book *Rolling Stones* as "It was not be found in the file, and no memorandum or date on the wrapper to show that it had ever been filed" (Henry, 1912). In this sentence, *file* occurs as a noun as in "a string or wire, on which papers and documents are strung for preservation and reference" (*oed.com*). This sentence also contains the verb *file* as in "to place (documents) on a file" (*oed.com*). To make sure *file* only appeared as a noun, POS tagging was used.

Another example of a loanword with a high frequency is *hunk*: 2.63 (1,066) (*COHA*) and 2.12 (1,566) (*COCA*). The Dutch loanword *hunk* is derived from Dutch *honk*, meaning “home base” (Van der Sijs, 2010, p. 245). However, the context function of *COCA* gives an example with a different meaning, which appeared on a website in 2012: "just a pack of pretzels […], a small hunk of cheese" (*yelp.com*). In this case, *hunk* is "a large piece cut off (e.g. from a loaf, cheese, etc.)" (*oed.com*).

For both cases, it does not mean that the loanwords have no Dutch history at all or are of an unrelated etymology. However, the words are used with a meaning different from their 'Dutch' meaning, which may explain why they have exceptionally high frequencies compared to loanwords which are used in the context of the 'Dutch' meaning. I will not focus on these high-frequency loanwords, as they are probably very productive and well-embedded in the language.

5.1.2. Exceptionally low frequencies

A couple of loanwords resulted in a frequency of 0.00 (0) in both *COHA* and *COCA*, namely *adstratum*, *barraclade*, and *rolliche*. Whilst *barraclade* occurs in the *OED*, Van der Sijs (2010) argues that it is now obsolete (p. 168). An explanation for this loanword not occurring in either corpora might be that, as it was a relic of the oldest Dutch settlers, it was practically obsolete by the nineteenth century (Van der Sijs, 2010, p. 169), so in those 80 years between the oldest documents of *COHA* and 1900, there was no source of record, at least in *COHA* and *COCA*.

Adstratum is used to "refer to linguistic elements that induced changes in a language that is otherwise dominant" (Van der Sijs, 2010, p. 276). This is a rather niche field and it

concerns jargon, which might be an explanation for the loanword not appearing in *COHA* or *COCA*.

According to Van der Sijs (2010), *rolliche* is only used regionally (p.138) and the *OED* mentions that the word is "now rare" (*oed.com*). These two arguments might be the reason why these loanwords had no frequency in both corpora.

5.2. Development over time

5.2.1. Ascending frequency over time

An ascending frequency of loanwords over time indicates that the frequency of a loanword increases over time or only came into use later. Ascending frequencies have multiple explanations. One explanation, related to the former increase, is that words are related to terms that gain popularity or expansion. An example of this explanation is *file*. This loanword's ascending frequency may be explained through the emergence of computers, where *file* is often used to refer to an electronic document stored on computers or other technological machines. See also *file* in section 5.3. below about the distribution of loanwords across text-types.

Another explanation for ascending frequencies is that the word is a term for something which did not even exist before. An interesting example of this is *Kuiper belt*. This loanword has no frequency in *COHA* until the 1990s. The first known Kuiper belt Object was discovered in 1992 (solarsystem.nasa.gov). This explains why there is no data of the loanword being used before this year and a sudden increase of the loanword after 1992, as people started talking about the object following its discovery.

5.2.2. Peaks in frequency

Some loanwords have a peak in frequency during a specific time period. This is illustrated in Appendices 4 and 5 by a sudden higher bar that is taller than its neighbouring bars. In both *COHA* and *COCA* you can click on the year(s) above the bars to see the distribution of each year (rather than all 10 years (*COHA*) or four years (*COCA*)). Naturally, all loanwords

have a peak in frequency, as they all have one highest frequency. In this section, however, I discuss the remarkable peaks, i.e., the sudden peaks in frequency, with no gradual increases.

A clear example of a sudden peak in frequency is *Fokker*, which seems to peak in the 1920s, with a frequency of 3.54 (91) in *COHA*. This is not random as in October 1919 the first flight of the Fokker took place[7] (fokker-history.com). As it was the first flight of a new aircraft manufacturer, which led the civil aviation market, it is likely that this loanword was used more during that time.

Another loanword with a sudden frequency peak is *hunk*. This loanword has a bar chart that peaks in the 1930s: 6.89 (191) in *COHA*. More specifically, *hunk* peaks in 1937 (28.75 (85) (*COHA*)). In this year, an American film was released, with a character named 'Hunk'. This might explain the sudden peak in frequency, as the word was more used in relation to the film at that time. See also *hunk* in section 5.3. below, which discusses the distribution of the loanwords across text-types.

Rip van Winkle also shows a peak, namely in the 1860s: 10.04 (170) in *COHA*. *Rip van Winkle* is a short story written by Washington Irving. This story was adapted for theatre and this adaptation opened in London in 1865 and on Broadway in 1866, hence the peak in frequency during that decade. When looking at the context in which *Rip van Winkle* is found in *COHA*, it is mostly used in the specific context of Irving's story, making the word less culturally embedded.

The first frequency of *superconducting/superconductor* is measured in 1935 in *COHA*. In this year, the brothers Fritz and Heinz London, proposed a theoretical model for superconductivity (London & London 1935). This explains the first overall frequency and a first peak. Another peak occurs in 1987, in which the Nobel Prize in Physics was awarded for a break-through discovery of superconductivity. A second explanation for the peak in frequency. Whilst these words are not etymologically Dutch, I have still counted them as

Dutch loans, because they are derived from the Dutch words *supergeleider* and *supergeleiding* (Van der Sijs, 2010, p. 276).

Pinkie has some incidental peaks in frequency, for example in 1920 (*COHA*) and 2010 (*COCA*). However, I could not find any relevant information that might explain the frequency of this loanword peaking in these years.

5.1.3. Random frequencies

Most loanwords have random frequencies, i.e., frequencies that do not follow a specific pattern, which is illustrated by bar charts that are disordered and unsystematic. Apart from a couple of loanwords that follow, for example, an ascending graph or a periodic graph, the majority of the words are more difficult to analyse and it is thus complex to form conclusions.

5.3. Distributions across text-types

Whilst the development of the frequency of loanwords over time can tell us a lot, so can the distribution of the frequency across text-types. This distribution provides more information about in which setting a loanword is more frequently used, such as formal versus informal, or written versus spoken language.

Both spellings of *bed*(-)*spread* show a peak of frequency in *COCA* in fiction: 4.02 (476) (*bedspread*) and 0.03 (3) (*bed-spread*). Whilst the latter is a relatively low frequency, especially compared to that of *bedspread*, it is the highest (and except one) only frequency for this spelling variant. Unfortunately, I could not find any relevant information related to this word in the context of fiction, which might explain the peak.

File peaks in webpages, with a frequency of 24.78 (3,079) (*COCA*). As discussed previously, this word is often used in the context of computers, hence the high frequency in this specific text-type.

Hunk occurs mostly in two text-types: tv shows and movies' subtitles (3.28 (420) in *COCA*) and fiction (4.30 (509) in *COCA*). As explained previously, Hunk was a character in the American film *Dead End*. This loanword also being a film character demonstrates the higher frequency in these two film-related text-types.

Skein peaks in the category of fiction (1.22 (144) in *COCA*). This may be explained by a book by Piers Anthony, which is called *With a Tangled Skein*.

The above-mentioned loanwords are words with a striking distribution across text-type, i.e., an interesting peak. Most loanwords, however, show no substantial differences in their distribution across text-types.

5.4. The emergence and disappearance of loanwords

Whilst the first data in *COHA* originate from 1820, many Dutch loanwords in American English did not yet exist at that time, or at least were not known through, for example, written texts or other sources. This is clearly reflected in the bar charts in *COHA*, where for a couple of decades there are no bar charts visible. This is the case for, amongst others, *banket* (first bar chart in the 1870s) and *bazoo* (first bar chart in the 1890s). As explained above, one explanation for loanwords having no frequency at first might be that the words have not been 'invented' yet.

On the other hand, some Dutch loanwords in American English have 'disappeared', at least according to *COCA*'s data. This is represented by no frequency or bar charts in the most recent years for multiple years at a time. This is the case for two loanwords: *bed-spread* (no bar chart from 2000-04 onwards) and *korfball* (no bar chart from 2000-04 onwards). Other loanwords might have no frequency for the last four (2015-19), e.g. *advocaat*, or eight (2010-19) years, e.g. *dobber*, but this can be incidental, as some loanwords, such as *bockey* and *kabouter* have these gaps of frequency as well, but the loanwords reappear after this time. This might also happen to *bed-spread* and *korfball*.

The disappearance of loanwords is also proven by the fact that Van der Sijs' (2010) glossary entails many loanwords that did not appear in the *OED*'s advanced search when

entering 1820 as the start date. This demonstrates that these Dutch loanwords had already disappeared by 1820.

5.5. Social meaning

As discussed in the literature review, loanwords carry social meaning (Blom & Gumperz 2000; Preston 2013; Zenner et al. 2019). Whilst it is possible to directly study the social meaning of loanwords, through, for example, attitude measurements, this is a different type of research from a corpus study. The former type includes factors such as community-based and speaker-based factors. The latter type, i.e., this corpus study, does not take these factors into consideration. However, through analysing the frequency of loanwords, I have attempted to investigate whether frequencies of loanwords can indicate any information about the social meaning of these words. The results have shown that the higher the frequency of a loanword, the more embedded a word is. However, the word does not always occur in its 'loanword' context, thus complicating the matter.

5.6. Further research

This research has attempted to investigate multiple aspects of Dutch loanwords in American English. Whilst my research has resulted in various conclusions, further research could be done to analyse other aspects, taking different concepts into consideration. Even though this research resulted in the majority of Dutch loanwords having a higher frequency in *COCA* than in *COHA*, it is important to note that other issues may play a role here. An example of such an issue is that *COCA* is a larger corpus than *COHA*: one billion words (*COCA*) versus 475 million words (*COCA*). Another issue is that *COCA* might contain greater diversity in text-types. Further research could take these issues into consideration and adjust some components accordingly.

More detailed research could be done in two ways. The first would be solely focusing on the development of the frequency of loanwords over time. For this, one would have to zoom in even more and analyse the frequency year by year, rather than in blocks of four or

10 years. Both *COHA* and *COCA* could still be used for this type of research, but each bar chart would have to be looked at individually. Another more precise investigation would be only looking at the distribution of the frequency of loanwords across text-types. This could be both formal versus informal text-types, but also within these text-types.

Whilst the *OED* considers the English language, it may not always be as effective regarding American English. There is still a difference between this type of English and (British) English. Not all Dutch loanwords in American English might have appeared in the *OED*, as some American English words might not even exist in the *OED*. A solution for this would be to use an American English-based corpus, such as an American dictionary or a dictionary of Americanisms.

Additional research could analyse the frequency of all words from Van der Sijs' (2010) glossary. Alternatively, further research could consider all the words with a Dutch etymology, which the *OED* provides.

6 Conclusion

Whilst speakers of Dutch and speakers of American English are relatively far away from each other geographically, their pasts and languages are connected. Due to the history of Dutch settlers and traders in northeastern America, the two languages came into contact. One of the results of this is Dutch loanwords in American English. Even though the Dutch did not stay in northeastern America, the influence they had on American English lasted and is still present today. This research aimed to investigate the frequency of Dutch loanwords in American English and concluded that the frequency of Dutch loanwords in American English increased over time.

The analysis for this research was performed based on previously done research by Schultz (2012, 2017, 2018, 2019) on the influence of multiple languages on English. Schultz's framework, using the *OED*'s advanced search engine was applied. Additionally, Van der Sijs' (2010) glossary of Dutch loanwords in American English was used. The overlap of words that resulted from the *OED*'s search and the words listed in Van der Sijs' (2010) glossary were entered in the *Corpus Of Historical American English* and the *Corpus Of Contemporary American English*. These two corpora provide frequency lists and bar charts, which allow for both an analysis of the frequency over time, as well as the frequency across text-types. The results were divided into semantic categories, which illustrate the different spheres of life of American English that were influenced by Dutch. This thesis also hinted at the social meaning attributed to loanwords.

Additional research could analyse the frequency of all Dutch loanwords in Van der Sijs' (2010) glossary, as well as all words that appear in the *OED* with a Dutch etymology. Additionally, further research could focus solely on the frequency of loanwords over time or across (and within) text types. American English-based corpora could be applied for further research

7 References

Backus, A. (2012). A usage-based approach to borrowability. *Tilburg Papers in Culture Studies* 27, 1-26.

Bedspread. (n.d.). In *Oxford English Dictionary*. Retrieved from https://oed.com.

Beer hall. (n.d.). In *Oxford English Dictionary*. Retrieved from https://oed.com.

Blom, J.P., & Gumperz, J. (2000). Social meaning in linguistic structure: Code-switching in Norway. *The Bilingualism Reader*, 111-136.

By the side of the road getaway lodging Harrisonburg (n.d.). Yelp. Retrieved May 26, from https://www.yelp.com/biz/by-the-side-of-the-road-getaway-lodging-harrisonburg.

Calque. (n.d.). In *Oxford English Dictionary*. Retrieved from https://oed.com.

Carpenter, W.H. (1908). Contributions to the vocabulary of English in America: Dutch remainders in New York State. *Modern Philology* 6(1), 53-68.

Davies, M. (2008). *The Corpus of Contemporary American English (COCA)*. https://www.english-corpora.org/coca/.

Davies, M. (2010). *The Corpus of Historical American English (COHA)*.https://www.english-corpora.org/coha/.

Dosker, N.H. (1880). De Nederlandsche taal in de Vereenigde Staten van Noord-Amerika [The Dutch language in the United States of North America]. *Noord en Zuid* 3, 45-48.

Durkin, P. (2014). Borrowed Words: A History of Loanwords in English. Detailed Record.

File. (n.d.). In *Oxford English Dictionary*. Retrieved from https://oed.com.

Fokker History. (n.d.). *Fokker F.II*. Fokker-History. https://www.fokker-history.com/f-ii.

Geeraerts, D. (2009). *Theories of Lexical Semantics*. Oxford University Press.

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

Henry, O. (1912). Rolling Stones.

Herbert, H.W. (1842). Sporty scenes and sundry sketches. *William Post* 1.

Hope we find our house in the same condition we left it. (2012, September 24). The Bayfield Bunch. Retrieved May 26, from http://thebayfieldbunch.com/2012/09/hope-we-find-our-house-in-same.html.

Hunk. (n.d.). In *Oxford English Dictionary*. Retrieved from https://oed.com.

Jacobs, J. (2009). *The colony of New Netherland: A Dutch settlement in seventeenth-century America*. Cornell University Press.

Kabouter. (n.d.). In *Oxford English Dictionary*. Retrieved from https://oed.com.

Laird, C.G. (1970). *Language in America*. World Publishing Company.

Language family. (n.d.). In *Oxford English Dictionary*. Retrieved from https://oed.com.

Llewellyn, E.C. (1936). *The influence of Low Dutch on the English vocabulary*. Oxford University Press.

Loanword. (n.d.). In *Oxford English Dictionary*. Retrieved from https://oed.com.

London, F., & London, H. (1935). The electromagnetic equations of the superconductor. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 149, 71-88.

National Aeronautics and Space Administration. (2018, December 14). *10 Things to know about the Kuiper Belt*. Solar System Exploration. https://solarsystem.nasa.gov/news/792/10-things-to-know-about-the-kuiper-belt/.

New World. (n.d.). In *Oxford English Dictionary*. Retrieved from https://oed.com.

Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics* 26(1), 47-104.

Preston, D. (2013). The influence of regard on language variation and change. *Journal of Pragmatics* 52, 93-104.

Provo. (n.d.). In *Oxford English Dictionary*. Retrieved from https://oed.com.

Schultz, J. (2012). Twentieth-century borrowings from French into English: An overview. *English Today* 28(2), 3-9.

Schultz, J. (2017). The impact of Japanese on the English lexicon since 1900. *Foreign Languages in Higher Education* 21(1), 41-58.

Schultz, J. (2017). The influence of German on the English lexicon in the twentieth century. *English Today* 33(1), 12-18.

Schultz, J. (2018). The impact of Spanish on English vocabulary since 1901. *English Today* 34(1), 26-31.

Schultz, J. (2019). The impact of Yiddish on the English language. *English Today* 35, 2-7.

Sociolinguistics. (n.d.). In *Oxford English Dictionary*. Retrieved from https://oed.com.

Te Winkel, J. (1896). Het Nederlandsch in Noord-Amerika en Zuid-Afrika [Dutch in North America and South Africa]. *Vragen van den dag*, 337-362.

Thomason, S.G., & Kaufman, T. (1988). *Language contact, creolization and genetic linguistics*. University of California Press.

Treffers-Daller, J. (2010). Borrowing. In M. Fried, J.O. Östman, & J. Verschueren (Eds), *Variation and change: Pragmatic perspectives. Handbook of pragmatics highlights* (pp. 17-35). John Benjamins Publishing Company.

Van der Donck, A.C. (1655). *Beschryvinge van Nieuw-Nederlant* [Description of New Netherland].

Van der Sijs, N. (1998). *Geleend en uitgeleend: Nederlandse woorden in andere talen & andersom* [Borrowed and lent: Dutch words in other languages & vice versa]. Uitgeverij Contact.

Van der Sijs, N. (2010). *Cookies, coleslaw and stoops: The influence of Dutch on the North American languages*. Amsterdam University Press.

Wales, K. (2014). A Dictionary of Stylistics. Routledge.

Walker, A., García, C., Cortés, Y., & Campbell-Kibler, K. (2014). Comparing social meanings across listener and speaker groups: The indexical field of Spanish /s/. *Language Variation and Change* 26(2), 169-189.

Zenner, E., Rosseel, L., & Calude, A.S. (2019). The social meaning potential of loanwords: Empirical explorations of lexical borrowing as expression of (social) identity. *Ampersand* 6, 1-4.

8 Appendices

Appendix 1

Loanword	***OED*** **(Oldest known quote)**	**Van der Sijs** **(Oldest known quote)**	**Difference (in years)**	**Information *OED***	**Information van der Sijs**	**Category (*OED*)**	**Category (vdS)**	**Category (own)**
adstratum	1939	1932	7	after Dutch *adstraat*	"In his inaugural speech in 1932, Dutch Romanist Marius Valkhoff was the first to use the linguistic term adstratum to refer to "linguistic elements that induced changes in a language that is otherwise dominant." He coined the word after the linguistic terms *substratum* and *superstratum*" (276)	Linguistics	Dutch loanwords that did not originate from immigrants	Language and linguistics
advocaat	1895	1945	50	late 19th cent. or earlier	Nineteenth or twentieth century (117)	Consu-mables: drink	Food, drink and stimulants - Liquor (117)	Food and drinks
banket	1886	1982	96	Dutch *banket* any of various kinds of sweet food	Nineteenth or twentieth century (117)	Consu-mables: food and cooking	Food, drink and stimulants - Confectio-nery (117)	Food and drinks
barraclade	1848	1848	0	Dutch *baare* bare, napless + *kleed* cloth.	From Dutch *baar kleed*, *bare kleden*, from *baar* ("naked, having no nap") and *kleed* ("cloth"); adopted in the seventeenth or eighteenth century and no longer extant (169)	/	Household effects and everyday implements	Household objects

bazoo	1877	1877	0	Dutch *bazuin* trumpet	Corruption of Dutch *bazuin* "trumpet"; borrowed in the nineteenth century and still commonly used as a slang word	Arts: music	Miscel-laneous	Arts
bedspread (OED) *bed-spread* (vdS)	1845	1845	0	Dutch *bed(de)sprei*	Perhaps after Dutch *bed-densprei*, *bedsprei* ("bed cover"), probably borrowed in the nineteenth century and still widely used (170)	/	Household effects and everyday implements	Household objects
beer hall (OED) *beer-hall* (vdS)	1837	1896	59	Dutch *bierhal*	From Dutch *bier-hal*, meaning "large bar serving beer in particular"; borrowed in the nineteenth century and still in use (234)	/	Trade	Miscel-laneous
bockey	1860	1848	12	Dutch *bokaal*	From Dutch *bakje*, *bakkie*, diminutive of *bak*, meaning "bowl"; adopted in the seventeenth or eighteenth century but now obsolete (171)	/	Household effects and everyday implements	Household objects
dingus	1866	1876	10	Dutch *dinges* item or person whose name the speaker does not know Dutch *ding* thing + *-es*, derivative suffix forming nouns denoting kinds of people, etc	From Dutch *dinges* meaning "what's his name"; borrowed in the nineteenth century and still in use (271)	/	Miscel-laneous	Language and linguistics

dobber	1836	1809	27	Dutch *dobber* float	From Dutch *dob-ber* meaning "float"; borrowed in the seventeenth or eigh- teenth century and still used regionally (175)	/	Household effects and everyday implements Fishing term (168)	Household objects Sport and leisure
dumbhead	1887	1887	0	Dutch *domkop*	After Dutch *dom-kop* meaning "stupid person," literally "dumb head"; probably borrowed in the nineteenth century and still com- monly used. (213)	/	Human traits and characteri-sations Characteri-sation (212)	Human charac-teristics
file	1851	1850	1	Dutch *feil*, variant or synonym of *dweil* floor-cloth	From dialectal Dutch *feil* "cloth used for wiping floors"; borrowed in the seventeenth or eighteenth century and still used regionally (176)	/	Household effects and everyday implements	Household objects
Fokker	1913	/	?	Name of A. H. G. *Fokker* (1890–1939), a Dutch engineer, the inventor.	The name of Dutch air-plane manufacturer Fokker ("he flew a Fokker") also made it into American English (276)	Transport: aviation	Dutch loanwords that did not originate from immigrants	Transport
fyke	1832	1832	0	Dutch *fuik*	From Dutch *fuik*, meaning "bow-net"; borrowed in the seventeenth or eigh- teenth century and still in use (176)	Sport and leisure: angling	Household effects and everyday implements Fishing term (168)	Household objects Sport and leisure
Hunk	1845	1848	3	Dutch *honk* goal, home, in a game	Derived from Dutch *honk*, meaning "home base"; adopted in the seventeenth century and still known (245)	/	Children's language Name for children's game (242)	Language and linguistics

kabouter (OED) *Kabouter* (vdS)	1961	/	?	Dutch	"a member of a Dutch group of political activists who promoted pacifism and anarchism; the kabouters – literally the 'dwarves' – were best known outside Holland for their 'white bicycles' scheme of the mid-1960s: white bicycles were left all over Amsterdam, free for anyone to use." (279)	Politics Religion and belief: mythology	/	Politics
korfball	1915	/	?	Dutch *korfbal*, < *korf* basket + *bal* ball	An older invention is that of *korfbal*, called korfball in American English. The name was made up by Dutch teacher Nico Broekhuijsen who founded the *Neder-landsche Korfbal Bond* (Dutch Korfball Association) in 1903. The sport became international in the 1970s and has been regulated and supervised by the International Korfball Federation since 1993. (277)	Sport and leisure: basketball	Sports	Sport and leisure
Kuiper belt	1988	/	?	the name of Gerard P. *Kuiper* (1905–73), Dutch-born U.S. astronomer, who suggested such a belt as the source of short-period comets (G. P. Kuiper in J. A. Hynek *Astrophysics* (1951) viii. 357 + belt	Kuiper belt (from Dutch *Kuipergordel)* (276)	Sciences: astronomy	Dutch loanwords that did not originate from immigrants	Science

logy	1859	1848	11	Dutch *log* heavy, dull	From Dutch *log*, "fat and plump, slow"; adopted in the seventeenth or eighteenth century, and still in use (215)	/	Human traits and characteri-sations Characteri-sation (212)	Human characteristics
lute	1875	1875	0	Dutch *loet*	From Dutch *loet*, meaning "tool consis-ting of a handle with a wide iron front piece used to dig or scrape"; adopted in the nineteenth century and still in use (177)	Manufac-turing and industry: brick-making	Household effects and everyday implements Tools (168)	Household objects
patroonship	1848	/	?	Dutch *patroonschap*	The *patroon-schappen*, named patroonships in American English (190)	/	Polity and citizens	Politics
pinkie	1840	1840	0	Dutch *pinkje* *pink* + *-je*, diminutive suffix	From Dutch *pinkje*, *pinkie*, a diminutive of *pink*, being the name of a particular type of vessel; adopted in the seventeenth century and still known (258)	Transport: nautical	Transport by sea and land Type of ship (256)	Transport
poppycock	1852	1852	0	Dutch *poppekak*, literally 'doll's excrement' (apparently only in the phrase *zo fijn als gemalen poppekak* showing excessive religious zeal, literally 'as fine as powdered doll's excrement	From Dutch *poppe-kak*, which literally means "doll's excrement"; adopted in the nineteenth century and still in use (273)	/	Miscel-laneous	Miscel-laneous
provo (OED) *Provo* (vdS)	1966	/	?	Dutch *provo* (also *Provo* as the name of the movement	In the 1960s, the image of the Provos or *the Dutch Provo movement* prevailed abroad. (278)	Politics	/	Politics

rijsttafel	1878	/	?	Dutch *rijsttafel*, literally ‘rice table’	rijsttafel, a comprehen-sive meal. The words were probably imported by (Indonesian) Dutchmen who, after Indonesia became independent in 1949, left the country and chose America as their new home rather than the Netherlands (118)	Consu-mables: food and cooking	Food, drink and stimulants Dish originating from Indonesia (118)	Food and drinks
Rip van Winkle (OED) *rip van winkle* (vdS)	1822	1833	11	*Rip Van Winkle*, the name of a character in Washington Irving's *Sketch Book* (1819–20), an amiable but lazy villager of Dutch de-scent, who falls asleep after a drinking party, sleeps through the American Revolutionary War, and awakes twenty years later in the United States of America	**rip van winkle** is a literary rather than a genuinely Dutch borrowing. (212) After the Dutch name of a character in a short story by Washington Irving; adopted in the nineteenth century and still in use. (217)	/	Human traits and characteri-sations Characteri-sation (212)	Human characteri-stics
rolliche	1830	1830	0	Dutch *rolletje*, literally ‘little roll’ (<*rol* roll + *-tje*, diminutive suffix),	From Dutch *rol-letje*, diminutive of *rol* (“roll”); adopted in the seventeenth or eigh- teenth century and still used regionally (138)	Consu-mables: food and cooking	Food, drink and stimulants Cold meats/meat dishes (117)	Food and drinks
sawbuck	1850	1851	1	Dutch *zaagbok*trestle, saw-horse	From Dutch *zaag-bok*, meaning “sawhorse”; probably borrowed in the nineteenth century and still in use (178)	/	Household effects and everyday implements Tools (168)	Household objects

scup	1848	/	?	Dutch *schoppen*, compare *schop*	scup, meaning "a swing" and to scup, meaning "to swing," from Dutch *schop* ("swing") and *schoppen* ("to swing") (242)	/	Children's language	Language and linguistics
skein	1837	1847	10	Dutch *scheen*	From Dutch *scheen* meaning "thin iron or wooden strip against wear"; probably adopted in the nineteenth century and still known regionally (179)	Crafts and trades: wood-working	Household effects and everyday implements Tools (168)	Household objects
smear-case	1829	1846	17	Dutch *smeerkaas*	From Dutch *smeer-kaas;* adop-ted in the nineteenth century and still in use (139)	Consu-mables: cheese	Food, drink and stimulants Cheese (117)	Food and drinks
snoop	1848	1848	0	Dutch *snoepen*	From Dutch *snoe-pen*, meaning "to eat sweets," whose literal meaning was probably adopted as early as the seventeenth or eighteenth century; still widely used but with different meanings and all sorts of derivatives (140)	/	Food, drink and stimulants Confec-tionery (117)	Food and drinks Language and linguistics
spearing	1884	1884	0	Dutch *spiering smelt*	From Dutch *spie-ring*, meaning "small silver white fish (Osmerus eperlanus)"; adopted in the seventeenth or eighteenth century and still used (166)	Sciences: fish	Flora and fauna Fish (144)	Flora and fauna

spook	1871	1801	70	Dutch *spoken*	From Dutch *spook* “ghost”; probably adopted in the seventeenth or eighteenth century and still in use, with new meanings (274)	/	Miscel-laneous	Language and linguistics
supercon-ducting	1913	1911	2	Dutch *supragelei-dend*	Various Dutch scientists coined a name for a new invention. For example, the terms superconductor and supercon-ducting are derived from the Dutch words *super-geleider* and *supergelei-ding*. The phenomenon was discovered and named in 1911 by Leiden-based scientist Heike Kamerlingh Onnes, who was awarded the Nobel Prize for it in 1913 (276)	Sciences: physics	Dutch loanwords that did not originate from immigrants	Science

supercon-ductor	1913	1911	2	Dutch *suprageleider*	Various Dutch scientists coined a name for a new invention. For example, the terms superconductor and supercon-ducting are derived from the Dutch words *super-geleider* and *supergelei-ding*. The phenomenon was discovered and named in 1911 by Leiden-based scientist Heike Kamerlingh Onnes, who was awarded the Nobel Prize for it in 1913 (276)	Sciences: physics	Dutch loanwords that did not originate from immigrants	Science
woonerf	1978	1964	14	Dutch *woonerf*, < *woon*-residential < *wonen* to live, reside) + *erf* ground, premises	In 1964, urban developer Niek de Boer built the first woonerf in the Dutch municipality of Emmen. The concept caught on and the nearly unpronounceabl e name was adopted, often accompanied by an explanation, such as “street for living,” “living street,” “living yard,” “residential yard,” “urban yard,” “living environment” or “home zone.” People even talk about *woon-erf streets* (277)	/	Dutch loanwords that did not originate from immigrants	Miscel-laneous

Appendix 2

Loanword	***COHA* total frequency**	***COHA* frequency per million words**	***COCA* total frequency**	***COCA* frequency per million words**	**Increase (>), decrease (<), same (=)**
adstratum	0	0	0	0	=
advocaat	0	0	5	0.01	>
banket	9	0.02	0	0	<
barraclade	0	0	0	0	=
bazoo	22	0.05	11	0.01	<
bedspread	463 (without '-') 12 (with '-')	1.14 0.03	689 (without '-') 3 (with '-')	0.93 0.01	> (without '-') < (with '-')
beer(-)hall	72 (without '-') 15 (with '-')	0.18 0.04	89 (without '-') 7 (with '-')	0.12 0.01	> (without '-') < (with '-')
bockey	0	0	2	0.00	>
dingus	54	0.13	115	0.16	>
dobber	1	0.00	1	0.00	=
dumbhead	18	0.04	9	0.01	<
file	1728	4.27	6612	8.93	>
Fokker	187	0.46	106	0.86	<
fyke	4	0.01	28	0.04	>
hunk	1066	2.63	1566	2.12	>
Kabouter/ Kabouter	0 (both spellings)	0	2 (both spellings)	0.00	>
korfball	0	0	4	0.01	>
Kuiper belt	115	0.28	453	0.61	>
logy	48	0.12	37	0.05	<
lute	1227	3.03	650	0.88	<
patroonship	2	0.00	0	0	<
pinkie	490	1.21	742	1.00	>
poppycock	116	0.29	143	0.19	>
provo/Provo	178 (both spellings)	0.44	457 (both spellings)	0.62	>
rijsttafel	5	0.01	2	0.00	<
rip van winkle/Rip van Winkle	423 (both spellings)	1.04	133 (both spellings)	0.18	<
rolliche	0	0	0	0	=

sawbuck	78	0.19	56	0.08	<
scup	4	0.01	41	0.06	>
skein	359	0.89	236	0.32	<
smear-case	0	0	0	0	=
snoop	320	0.79	1287	1.74	>
spearing	168	0.41	234	0.32	>
spook	470	1.16	927	1.25	>
supercon-ducting	103	0.25	567	0.77	>
supercon-ductor	49	0.12	196	0.26	>
woonerf	0	0	0	0	=

Appendix 3

Semantic categories

The *Oxford English Dictionary*

1. Arts
    1. Music
2. Consumables
    1. Cheese
    2. Drink
    3. Food and cooking
3. Crafts and trades
    1. Woodworking
4. Linguistics
5. Manufacturing and industry
    1. Brick-making

6. Politics
7. Religion and belief
    1. Mythology
8. Sciences
    1. Astronomy
    2. Fish
    3. Physics
9. Sport and leisure
    1. Angling
    2. Basketball
10. Transport
    1. Aviation
    2. Nautical

Llewellyn (1936)

1. Intercourse through Trade between Britain and the Low Dutch Countries (p. 32)
2. Intercourse between English and Low Dutch Fishermen (p. 89)
3. The Low Dutch and the Manufacture of Cloth (p. 115)
4. The Influence of Low Dutch on the Technical Vocabulary of Various Crafts and Manufactures (p. 131)
5. The influence of Low Dutch on the English Vocabulary of Science, Literature, and Art (p. 143)
6. English and Dutch Intercourse in North America (p. 173)
7. Unclassified Borrowings from Low Dutch (p. 186)

Van der Sijs (2010)

1. Food, drink, and stimulants (p. 117)
2. Flora and fauna (p. 144)
3. Household effects and everyday implements (p. 167)
4. Polity/Policy and citizens (p. 180)

5. The American landscape (p. 201)
6. Human traits and characterisations (p. 212)
7. Religion and religious festivals (p. 219)
8. In and around the house (p. 226)
9. Trade (p. 232)
10. Money and units of measure (p. 237)
11. Children's language (p. 242)
12. Transport by sea and land (p. 256)
13. Clothing (p. 264)
14. Miscellaneous (p. 270)
15. Dutch loanwords that did not originate from immigrants (p. 276)

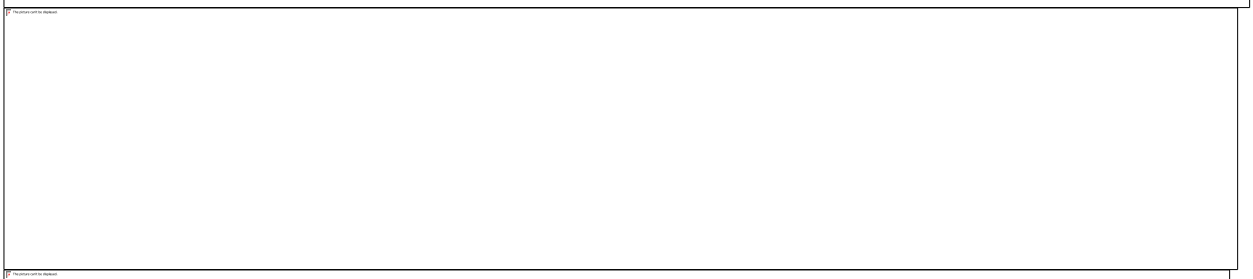
Schultz (2012, 2017, 2018, 2019)

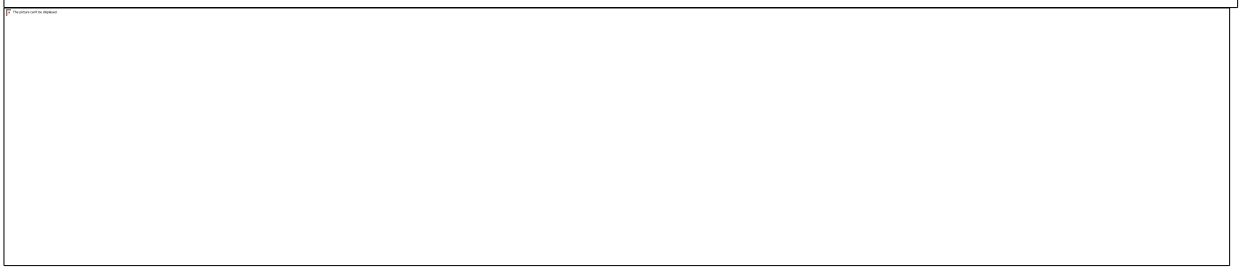
1. Anthropology (2012)
2. Metapsychics and parapsychology (2012)
3. Archaeology (2012, 2018)
4. Miscellaneous (2012)
5. Technology (2012, 2017)
6. La Francophonie (2012)
7. Fashion and lifestyle (2012)
8. Entertainment and leisure activities (2012, 2019)
9. Mathematics and the humanities (2012, 2017, 2018)
10. People and everyday life (2012, 2017, 2018, 2019)
11. Civilization and politics (2012, 2017, 2018)
12. Gastronomy (2012, 2017, 2018, 2019)
13. Fine arts and crafts (2012, 2017, 2018)
14. The natural sciences (2012, 2017, 2018)
15. Leisure and pleasure (2017, 2018)
16. The material arts (2017)
17. Culture and history (2017)

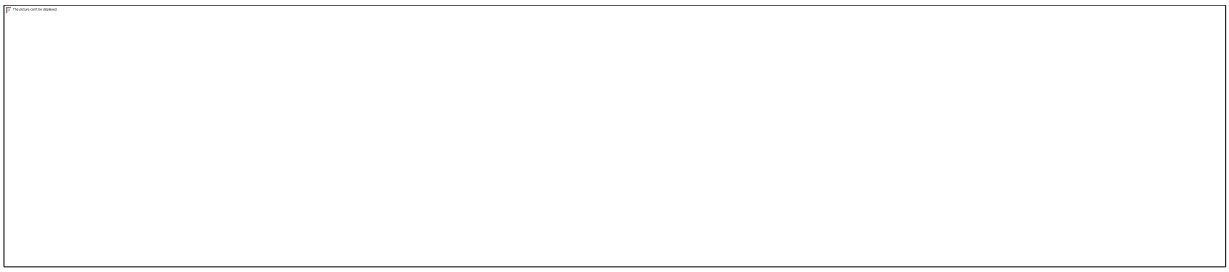
18. Science and technology (2017)
19. Language and linguistics (2019)
20. Criminality (2019)
21. Faith and religion (2019)
22. Technology and electronics (2018)

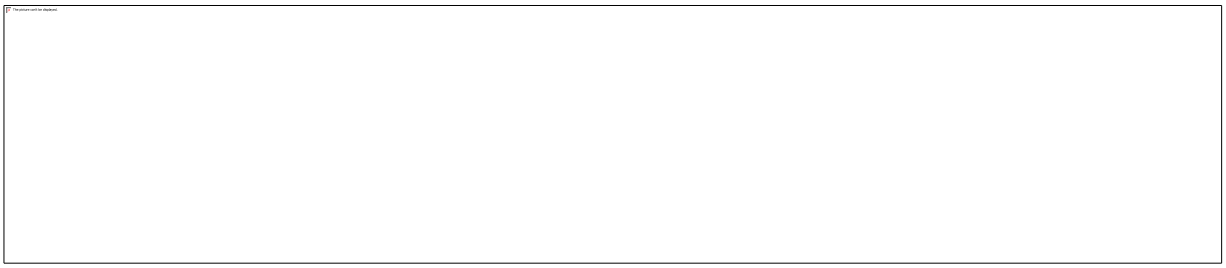
Appendix 4

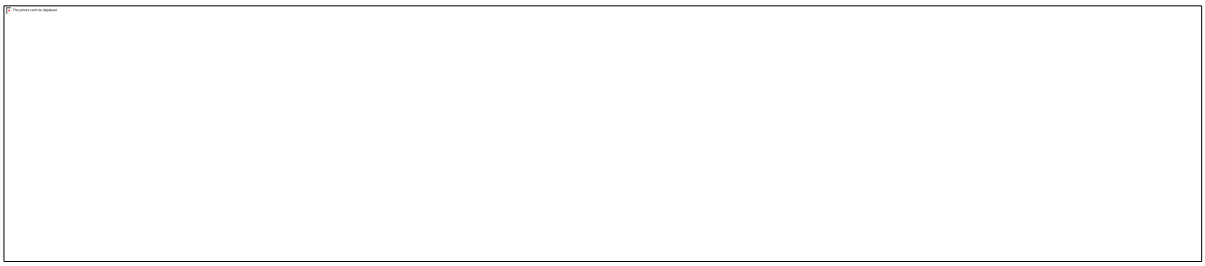
Screenshots from bar charts from the *Corpus of Historical American English*

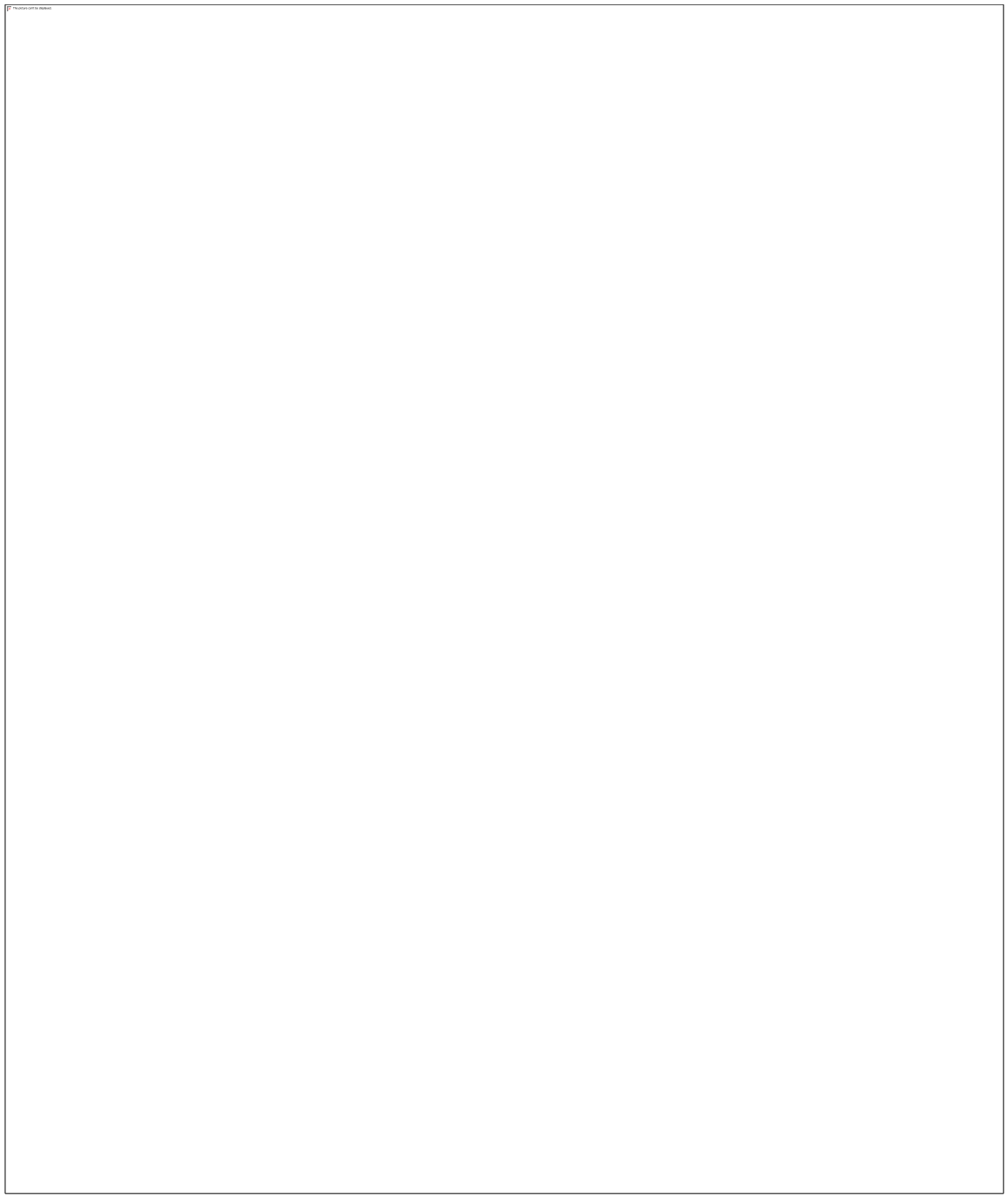





The picture can't be displayed.


The picture can't be displayed.


The picture can't be displayed.


The picture can't be displayed.


The picture can't be displayed.


The picture can be displayed.

Appendix 5

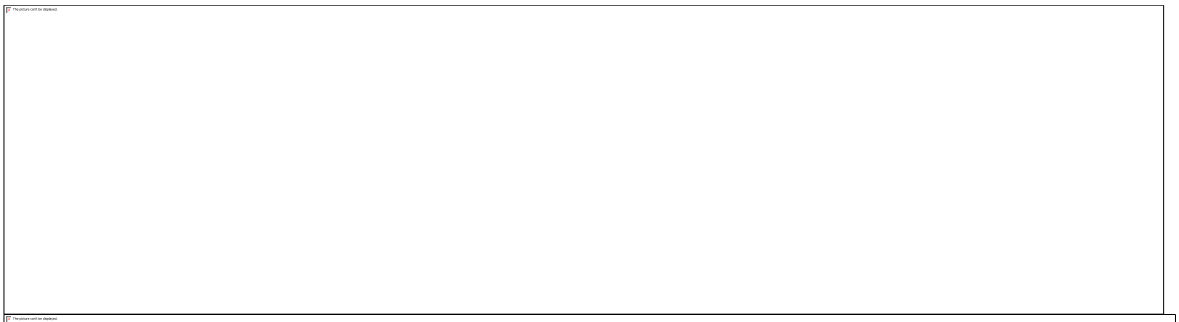
Screenshots from bar charts from the *Corpus of Contemporary American English*

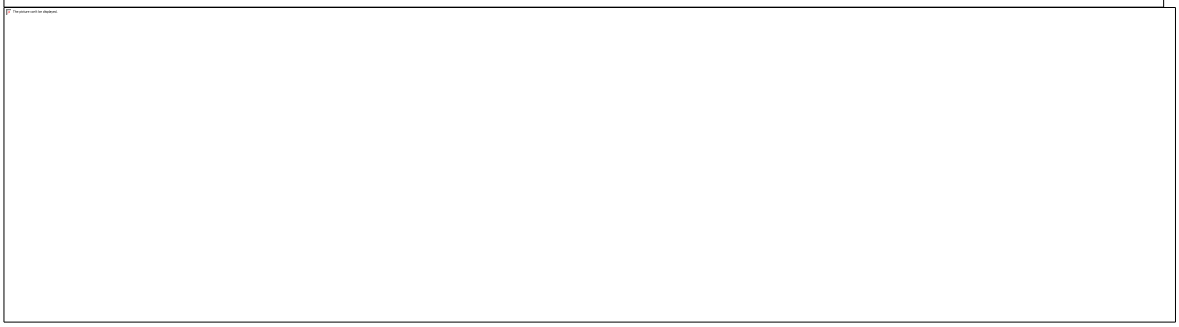
The picture can't be displayed.

The picture can't be displayed.

The picture can't be displayed.


The picture can't be displayed.


The picture can't be displayed.


The picture can't be displayed.