Use of English words in the Dutch language

Patrick Jans, S4525205
January 16, 2024
Link to Github page:

https://github.com/PatrickM-J/Intro-to-RM/tree/main

Abstract

Today, a lot of Dutch speakers tend to use or replace Dutch words with their English versions. This seems to happen more frequently especially under younger age groups. Due to the increasing presence of English in the Dutch language, we analyzed if there is a difference in the frequency of English words used in two different age groups (Ages 16-30 and 30+). We accomplished this by first identifying all English words in a corpus from tweets by Dutch users. Then, we calculated the percentage of total words that were English. We found that approximately 34% of words were English in the age group 16-30, while only 18% of words were English for the age group 30+.

1 Introduction

The past few decades the Dutch language is being influenced more and more by the English language. Not only have certain English words been added to the Dutch language ('loanwords' or 'leenwoorden' in Dutch), English is also being used more frequently in other aspects such as education and business. For instance, in the year 2024/25 47% of university bachelor degrees will be taught in English in the Netherlands (Keuzegids, 2023). Furthermore, the use of English is also very common in different kinds of media. On social media for instance, users (especially younger age groups) tend to often use English words such as 'nice' or 'chill'. Moreover, in most popular video games, tv shows and films the main language is English, with mostly only subtitles provided.

In order to try and understand the prevalence of English words in the Dutch language, this study wants to analyze if there is a clear difference in the frequency of English words used by Dutch people in different age groups. In order to do this, we will be analyzing tweets placed by Dutch users. By counting the amount of English words used, we can calculate the percentage of English words present. Hopefully this will help us better understand how language can differ between age groups. Furthermore, it could give us more insight into how a language can evolve over time due to the influence of another language.

For this research, we will be analyzing data from two different age groups: ages 16-30 and ages 30 and above. Our expectation is that the younger age group will use English words in their language more often. People from this age group tend to use social media more often, since they come into contact with English frequently. Additionally, younger people also use other kinds of media where English is the standard language, such as video games and films and series.

2 Related Work

Earlier studies have already investigated the use of English words in the Dutch language. In a study by Gerritsen et al. (2000), they investigated how often Dutch television commercials contained English and whether consumers understood and appreciated the English used. In our case, we are mostly interested in the frequency of English used in commercials. They found that one-third of the commercials were partly or completely in English, with most of the commercials consisting for one-fifth of English text. These commercials date back to 1996, but although the data is out-dated, we can still use it as a comparison to show the

development of English in Dutch language.

In another similar research by Smakman et al. (2009), they, analyze the extent of English words used in Dutch radio commercials and the level of understanding and appreciation of using English in commercials. They found that of a total of 84 commercials, 33 (39%) contained at least one English words, with the average percentage of English words in a commercial being only 5%. A given explanation for the relatively low frequency of English words is due to short run-time of radio commercials compared to television commercials, which may make English used on radio commercials harder to understand.

Finally, in an article by van der Sijs (2023) for the scientific magazine OnzeTaal, the amount of English loanwords were analyzed of the first five pages of an edition of the Dutch newspaper NRC. It was found that there were 150 distinct English loanwords which were used 281 times in total. Similar samples were taken in 1994 and 2012, and compared to 1994, both values increased by 300% as seen in figure 1. Furthermore, they also found that the age of the journalist of an article showed a clear relation with the amount of English words used: the younger the journalist, the more English loan words were used.
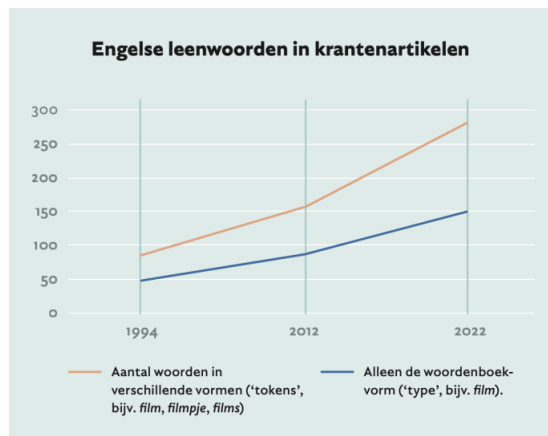

Figure 1: From van der Sijs (2023)

3 Data

To analyze the frequency of English words used by Dutch speakers, we will be examining tweets from Dutch users. We cannot simply use existing tweets from Twitter itself, since information such as gender and age can sometimes be hard to determine. Therefore, we will be using a corpus with tweets from Dutch users where information such as their age is already known, so that we will be able to differentiate between the two age groups.

We will be computing the frequency of English words used as a percentage of the total amount of words. In order to determine whether a word is either English or Dutch, we will be using a similar method as done by Gerritsen et al. (2000) and Smakman et al. (2009). In these two studies, they used the at the time most recent version of the Dutch dictionary Van Dale to determine the origin of words. If a word we encounter in a tweet is found in the most recent edition of the Van Dale (in our case the 16th edition published in 2022), we will consider it a Dutch word.

Unfortunately, there are some exceptional cases which our method will incorrectly determine to be either English or Dutch. Certain words in Dutch and English are interlingual homographs: words with identical forms in two different languages but with different meanings. A good example for English and Dutch would be the word 'of', which means 'or' in Dutch but is the English version of the Dutch word 'van'. In order to prevent this mistake, we will not be counting these interlingual homographs.

Furthermore, due to the nature of twitter, some words will be misspelled or abbreviated. In the first case, we will simply ignore these cases. It would be to complicated to take into account all possible spelling mistakes made by users. Additionally, some spelling mistakes could have multiple possible words which they were supposed to be. For the abbreviations, we could include a list of a number of very popular abbreviations such as BTW (By The Way) and DM (Direct Message). However, it does complicate how we should count these words. For instance, BTW is technically an abbreviation for three separate words. Should we count it as just the single one, or as all three? Therefore, we will also ignore these special cases.

Table 1 provides a summary of the data that will be used in this study.

Age:	Percentage of English words used
16-30	
30+	

Table 1: Contingency table for percentage of English words used on Twitter

4 Predicted Results

We expect that the younger age group (ages 16-30) will use English words more often than the older age group (ages 30+). Our expectation is that the percentage from ages 16-30 will be approximately 34%. For ages 30+ we expect it will be approximately 18%. In general as age increases, social media activity decreases (Hruska and Maresova (2020)) and young adults from ages 18 till 29 are also most likely to use social media as well (Perrin (2015)). On social media they come into contact with English more frequently, since a large part of all content is in English. Therefore we expect due to more interaction with the English language, that they will use it in their own language more often. Furthermore, it has already been shown by van der Sijs (2023) that the age of a journalist shows a clear relation with the use of English loanwords used. The younger the journalist is, the higher the amount of English loanwords used.

Table 2 summarizes an example of the expected results for our study:

Age:	Percentage of English words used
16-30	34%
30+	18%

Table 2: Contingency table for percentage of English words used on Twitter as a percentage of the total amount of words

4.1 Discussion

If the results are similar to our expectations, this could be an interesting observation on how the Dutch language is slowly incorporating more and more English words. Recently this process, also called 'Anglicisation' or 'Verengelsing' in Dutch, has gotten media and political attention as well.

Some people fear that the growing influence of the English language will eventually cause the Dutch language to lose its importance or even vanish. The Anglicisation on Dutch universities has even become a political topic, with plans to limit the amount of studies given in English and encouraging international students to learn Dutch as well.

However, incorporating words from other languages into the Dutch language is hardly a recent phenomenon. The Dutch language has been adopting loanwords from different languages such as French and Latin for centuries (van der Sijs (2023)). Other languages have also done the same: we even have adopted loanwords from English that originate from French such as 'alert' or 'veggie'. Due to the importance of the English language on a global level, it has become more prevalent in the Dutch language.

5 Conclusion

The aim of this study was to find out if there is a difference in the frequency of English words used by two different age groups. We expect that the younger age group (ages 16-30) will use English words more often. People from this age group in general come into contact more often with English through media such as social media platforms, films, television and the internet. Moreover, younger generations in the Netherlands are getting taught English in school earlier than before, with English now being a mandatory subject in Dutch primary schools.

However, it is important to note the context surrounding our research. Twitter is a social media platform where English words and terms may be encountered more easily due to most content being in English. This may cause people to use English words on Twitter more often than they would in their everyday life.

Additionally, due to limitations, we were only able to focus on two age groups. The age group 16-30 contains teenagers still in high school, but also people at the end of their twenties who might already have been working for a couple of years. Moreover, the age group 30+ is even more diverse. It is expected that by only examining these two age groups, we generalize these statistics for a large age group which contains many stages of

life. By analyzing the frequency of English words used for more precise age groups, we can better understand the influence of the English language for different ages instead of just the two age groups we analyzed in this study.

Furthermore, as described in the Data section, it is very difficult to categorize some words as either Dutch or English. For further research, looking for a method to better capture all words used could improve the categorizing of English and Dutch words.

References

Gerritsen, M., H. Korzilius, F. Van Meurs, and I. Gijsbers (2000). English in dutch commercials: Not understood and not appreciated. *Journal of advertising research 40*(4), 17–31.

Hruska, J. and P. Maresova (2020). Use of social media platforms among adults in the united states—behavior on social media. *Societies 10*(1), 27.

Keuzegids (2023, Nov). Internationalisering in het wo neemt verder toe.

Perrin, A. (2015). Social media usage. *Pew research center 125*, 52–68.

Smakman, D., H. Korzilius, F. Van Meurs, and E. Van Neerven (2009). English words and phrases in radio commercials in the netherlands: Their use and effects. *ESP across Cultures 6*, 107–128.

van der Sijs, N. (2023). Een 'boom' aan engelse leenwoorden. *OnzeTaal*.