



RU:ts

STUDENT LINGUISTICS JOURNAL
Radboud University Nijmegen

EDITION 2,
FEBRUARY 2021



INTRODUCTION

After months of work by authors, reviewers, and the editorial team, we finally have a second edition!

This issue contains excellent work by Radboud students, both bachelor's and master's students, about their research projects in all fields concerning language and linguistics. Topics that are covered in the articles are colour terms across Europe, language and homosexuality, the use of tense in the series 'How I met your mother', classification of nouns in Avatime, and language relativity.

The papers are also available online for free, as they are published under Open Access Creative Commons Licence: the authors receive the rights to their papers, and can (re)publish them anywhere as long as they mention RU:ts in some way!

As new editors-in-chief we would like to thank our predecessors, Iris Faber and Myrthe Reuver, for setting up this journal. After publishing the first edition they started working on the second edition and we took over halfway through. The collaboration between the two boards therefore makes this edition extra special. In addition, we would like to thank the rest of the old board for all their hard work.

We also want to give special thanks to the teachers who helped us during the process of publishing the second edition, especially dr. Nelleke Oostdijk and prof. dr. Helen de Hoop. Finally, we would also like to thank all the reviewers and authors who made this second edition possible!

Best wishes,
Jasmijn Vestjens & Michelle Suijkerbuijk
Editors-in-Chief

COLOFON

Editors-in-Chief

Jasmijn Vestjens
Michelle Suijkerbuijk

Editors

Lynn de Rijk
Marieke van den Akker

Lay-out editor

Elsa Opheij

Treasurer

Simon Claassen

PR & Social Media

Abdellah Elouatiq

SPECIAL THANKS TO:

The old board

Dr. Nelleke Oostdijk

Prof. dr. Helen de Hoop

Department of
Language and
Communication

And, most importantly,
our authors and
reviewers!





**COLOR TERMS ACROSS EUROPE: AN INVESTIGATION OF THE
INTERACTION BETWEEN COLOR TERMS AND EUROPEAN LANGUAGE
FAMILIES**

S.A. (SIMON) CLAASSEN

P. 3

POLARI: LANGUAGE AS A MEANS OF CONCEALMENT

R. (ROEL) STROO & A.A.K. (BRAM) PEUTE

P. 22

**WAAROM WOORDEN IN KLASSEN PASSEN: SEMANTIEK IN HET
NOMINALE CLASSIFICATIESYSTEEM VAN HET AVATIME**

S.E.M. (SAM) THEUNISSEN

P. 32

**LANGUAGE RELATIVITY RE-VISITED: PERCEPTION OF BLUE AND GREEN
IN GREEK, IRISH, AND GERMAN**

M.C.A. (MIGNON) VAN VORSELEN

P. 53

**TED TALKING TO HIS KIDS: THE PRESENT PERFECT IN ENGLISH AND
DUTCH IN 'HOW I MET YOUR MOTHER'**

F.J.G. (VEERLE) WILMS

P. 62



Color terms across Europe

An investigation of the interaction between color terms and European language families

S.A. (Simon) Claassen

Research Master *Linguistics and Communication Sciences*, Radboud University Nijmegen

Abstract: Languages show variation in the colors that they have a distinct color term for, with some languages only distinguishing two colors and some having distinct terms for up to eleven different colors. These appear to follow a hierarchy, where colors at a higher place in the hierarchy will have color terms in a certain language before those at a lower place. Although many studies have been conducted so far that focused on the different color terms found in various languages, there have as of yet been no studies that have attempted to find patterns in the distributions of these terms across related languages. Therefore, this paper investigated what the etymological origins of the color terms of European languages are. Additionally, it examined whether the forms of color terms found in languages are categorically more similar when these languages are from the same language family or geographical area. Finally, it investigated whether the number of colors distinguished within a language can be related to the language family this language belongs to. For this purpose, the color terms of fifty European languages, some related and some unrelated, were collected and compared. Firstly, it was found that languages usually acquire color terms by inheriting them from their proto-language, especially for the upper colors in the hierarchy, whereas other strategies to acquire color terms, such as borrowing or morphological derivation, are generally only used for the lower colors of the hierarchy. This supports the aforementioned color hierarchy, as it implies that the upper colors of the hierarchy are more semantically stable across different stages of language development. It was also found that the color terms found in a language are influenced to a greater degree by its language family than its geographical location. Finally, no link was found between the numbers of color terms found in closely related languages.

Key words: color terms, color hierarchy, language family, Indo-European linguistics, historical linguistics

1. Introduction

As human beings, we constantly experience the world around us through our senses. Arguably our most important sense, vision, allows us to perceive objects in our surroundings through the light reflected by these objects (Livingstone, 2002). The wavelengths of the electromagnetic spectrum that are reflected or absorbed by objects determine the colors these objects appear to have to our eyes. The color spectrum of visible light encompasses wavelengths ranging from approximately 400 to 700 nm, with the lower wavelengths of the spectrum appearing to our eye as violet and the higher ones appearing as red (Buser & Imbert, 1992).

Although the color spectrum forms a continuous unit comprising numerous different hues gradually transitioning from one to another, languages require people to categorically divide the spectrum into distinct colors for the sake of naming these colors with respective color terms. Languages have been shown to display considerable variation with regard to the colors they distinguish (Kay, Berlin, Maffi, & Merrifield, 1997). Most European languages linguistically distinguish a wide variety of colors and specific shades, whereas there are also languages, such as Tsimané, that reportedly have distinct terms for only two or three colors (Gibson et al., 2017), with additional claims that there are even languages without any color terms, such as Pirahã (Regier, Kay, & Khetarpal, 2009).

Generally, however, languages do not appear to distinguish more than eleven distinct color categories. These color terms appear to follow a hierarchy, where colors higher up in the hierarchy must be present in languages before colors further down in the hierarchy (Berlin & Kay, 1991). This color hierarchy is shown in Figure 1 below. The semantic boundaries of color terms also appear to vary from language to language. For instance, whereas most European languages have two separate words for the colors green and blue, many languages group these two colors together under a single term (Davies & Corbett, 1997).

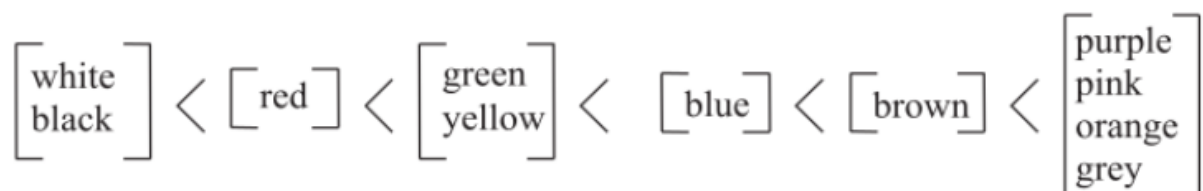


Figure 1. The color hierarchy (Berlin & Kay, 1991).

These observations have raised several questions related to cognition. One could wonder, for instance, whether the color terminology of a language shapes or influences its speakers' perception of certain colors or whether people are born with certain color

concepts already in place. It appears that language does indeed influence color perception to some degree, but there also appears to be evidence for the existence of universally present color categories, which are often stated to be black, white, red, green, yellow and blue (Regier & Kay, 2009).

Over the years, there have been multiple studies that attempted to report on the different color terms of numerous languages. The biggest study to do so was the World Color Survey, which gathered data on color terminology from 110 languages (Kay, Berlin, Maffi, Merrifield, & Cook, 2009). From studies such as this one, it becomes evident that the languages of the world are very diverse with regard to the form, meaning and categorization of their color terms. Languages can have both morphologically simple and complex forms for color terms, have just a few or quite a lot of color terms, distinguish certain colors or group multiple colors under a single term and place the boundaries between certain colors in different places.

It would be interesting to relate the dispersion of color terms across the languages of the world to certain topics in linguistic typology and historical linguistics. The continent of Europe is a typologically interesting geographical area. Almost all of the languages spoken on the continent, both national and regional, belong to the same linguistic family, the Indo-European languages. Languages belonging to the same subfamily within this language family can be clearly seen to share more features with each other than with other branches within the family, quite often because certain words and grammatical structures in both languages derive from their common ancestor language, often called a proto-language (Renfrew, 1989).

One can thus imagine that the color terms found in languages and, by extension, the demarcations of the color categories these color terms refer to might be more similar to each other the more closely related these languages are. Another factor could be geographical distribution (Epps & Huehnergard, 2013). As languages that are spoken in the same geographical area and have been in contact with each other over time tend to adopt words from each other, it is also possible that languages show similarities in their color terms with the languages in their direct vicinity.

Analyzing a sample of fifty European languages, I will examine three aspects of the relations of color terminology between these languages, corresponding to my three main research questions. I will investigate what the etymological origins of the color terms of European languages are, whether the forms of color terms found in languages are categorically more similar when these languages are from the same language family or geographical area and whether the number of colors distinguished within a language can be related to the language family this language belongs to.

As for the first question, I expect that most color terms for colors high up in the color hierarchy cannot be traced back to etymologically originate in a so-called source-based term, which is a color term that is clearly named after an object of this color (de Valk, Wnuk, Huisman, & Majid, 2017). Additionally, I expect that some color terms, especially those for the lower colors of the color hierarchy, will have a clearly source-based origin or be derived from other color terms. I also expect that color terms that cannot be traced back to originate in source-based terms will mostly, though not necessarily always, retain the same semantic meaning and thus refer to the same color throughout different proto-stages of languages.

As for the second question, I hypothesize that languages within the same families are more likely to have the same color terms and that languages within the same geographical area are also more likely to share certain color terms, although presumably to a lesser degree than languages within the same family. Finally, as for the final question, I hypothesize that there will be few differences in the number of colors distinguished by European languages, as I expect that most European languages distinguish all of the eleven colors in the sample, or else ten or nine at least, and that, if there are differences in the number of colors distinguished, these will likely pattern with language families.

2. Method

In order to test the hypotheses mentioned above, I collected language data on color terms from a sample of fifty European languages to form a relatively representative sample. To this end, the selected languages are roughly equally dispersed across Europe geographically and across the various language families and subfamilies of Europe. The majority of languages in the sample are Indo-European languages belonging to one of the six branches of this family spoken in Europe, along with three more languages from the Uralic family as well as the Basque and Maltese languages.

All languages in the sample are proportionally distributed across the primary and secondary levels of their respective language families. The three biggest subgroups in the sample, the Germanic, Italic and Balto-Slavic languages, are further divided into the West and North Germanic, the Western, Central, Southern and Eastern Romance and the West, South and East Slavic and Baltic subfamilies, respectively. Of course, many of these language families can be further divided into even smaller subgroups, but for practical reasons, this classification was used. The languages included in the

sample are shown in Table 1 below. In Figure 2, one can see the geographical distribution of these languages throughout Europe.

Table 1

Languages included in the sample and the (sub)families they belong to.

Language	Family	Language	Family
English	Indo-European; Germanic (West)	Manx	Indo-European; Celtic
Frisian	Indo-European; Germanic (West)	Irish	Indo-European; Celtic
Dutch	Indo-European; Germanic (West)	Gaelic	Indo-European; Celtic
German	Indo-European; Germanic (West)	Welsh	Indo-European; Celtic
Danish	Indo-European; Germanic (North)	Cornish	Indo-European; Celtic
Swedish	Indo-European; Germanic (North)	Breton	Indo-European; Celtic
Norwegian	Indo-European; Germanic (North)	Greek	Indo-European; Hellenic
Icelandic	Indo-European; Germanic (North)	Albanian	Indo-European; Albanian
Faroese	Indo-European; Germanic (North)	Polish	Indo-European; Balto-Slavic (West)
Romansh	Indo-European; Italic (West)	Czech	Indo-European; Balto-Slavic (West)
Ladin	Indo-European; Italic (West)	Slovak	Indo-European; Balto-Slavic (West)
Friulian	Indo-European; Italic (West)	Slovene	Indo-European; Balto-Slavic (South)
Lombard	Indo-European; Italic (West)	Serbian	Indo-European; Balto-Slavic (South)
Piedmontese	Indo-European; Italic (West)	Bulgarian	Indo-European; Balto-Slavic (South)
French	Indo-European; Italic (West)	Macedonian	Indo-European; Balto-Slavic (South)
Occitan	Indo-European; Italic (West)	Russian	Indo-European; Balto-Slavic (East)
Catalan	Indo-European; Italic (West)	Ukrainian	Indo-European; Balto-Slavic (East)
Aragonese	Indo-European; Italic (West)	Belarusian	Indo-European; Balto-Slavic (East)
Asturian	Indo-European; Italic (West)	Lithuanian	Indo-European; Balto-Slavic (Baltic)
Spanish	Indo-European; Italic (West)	Latvian	Indo-European; Balto-Slavic (Baltic)
Portuguese	Indo-European; Italic (West)	Estonian	Uralic; Finnic
Italian	Indo-European; Italic (Central)	Finnish	Uralic; Finnic
Sicilian	Indo-European; Italic (Central)	Hungarian	Uralic; Ugric

Sardinian	Indo-European; Italic (South)	Basque	Vasconic
Romanian	Indo-European; Italic (East)	Maltese	Afroasiatic; Semitic



Figure 2. Geographical distribution of languages included in the sample.

For each of these languages, appropriate linguistic descriptions or lexicons were collected and terms for the colors red, orange, yellow, green, blue, purple, pink, white, gray, black and brown were retrieved from these sources. Usually, it was quite straightforward and easy to see what the main term used for a color was, but if there were multiple terms in common usage for a certain color in a language, the term that appeared to be more frequent or more basic in meaning or usage was chosen. This, however, was not always obvious. For instance, the German word *purpurn* appears to be a pretty basic term for the color purple, but it is not really used by German speakers anymore, with *violett* being the preferred term.

3. Results

The sample of color terms retrieved from the literature can be found in the Appendix, as well as the sources used for each language. As can be seen, each of the fifty languages in the sample has a color term of some sort for each of the eleven color categories. For the six highest colors of the color hierarchy, black, white, red, yellow, green and blue, many languages have a color term they inherited from their respective proto-language. Most of them are monomorphemic and cannot be traced back to have source-based origins.

Almost all of the terms for these colors also retained their original meaning and thus consistently referred to the same color through different stages of language development. For instance, the Dutch word *rood* “red” descends from Proto-Germanic **raudaz* “red”, which in turn descends from the Proto-Indo-European root **h₁rewdʰ-* “red” (Kroonen, 2013). Similarly, the Spanish word *verde* “green” descends from Proto-Romance/Vulgar Latin *viridis* “green” and ultimately from the Latin word *viridis* “green” (de Vaan, 2008). As can be seen, these two terms were passed down from two proto-languages to their descendants and retained their meaning over time.

There were only a few instances where the most commonly used term in a language for one of the six highest colors of the hierarchy originated in a source-based term or a term for a different color. One of the most notable examples is the Slovene word *rumena* “yellow”, which derives from Proto-Slavic **ruměnbъ* “reddish, rosy”, also from Proto-Indo-European **h₁rewdʰ-* “red” (Derksen, 2007). Slovene does have another word for yellow, *žolta*, which is a cognate of the word for yellow in the other Slavic languages, but this word appears to be very rare, apparently having been largely displaced by *rumena*.

Another interesting example of such a semantic shift concerns Lithuanian *mėlyna* “blue” and Latvian *melns* “black”, which both derive from Proto-Baltic **mel(n)-* “black, blue” and ultimately from Proto-Indo-European **mel-* “dark, red” (Derksen, 2014). As can be seen, these two terms for colors at the top of the hierarchy both derive from a term originally referring to a different color, possibly through the meaning of “dark” of this latter term. Not only have these two color terms diverged in meaning from their origin, they have also diverged from each other over time.

If languages lack monomorphemic or etymologically stable terms for certain colors, this happens almost exclusively in the lower colors of the hierarchy. Languages occasionally derive some of these colors directly from other color terms present in the language. For instance, the Danish word *lyserød* “pink” and the Faroese word *ljósareyður* “pink” are both derived from the respective words for red in these languages, both literally translating to “light red”. Similarly, the Manx words *jiarg-*

bwee “orange”, *jiarg-gorrym* “purple” and *jiarg-bane* “pink” literally translate to “red-yellow”, “red-blue” and “red-white”, respectively.

Languages also appear to resort to source-based terms increasingly for the lower colors of the hierarchy, although the boundary between actual source-based terms and lexicalized color terms has proven to be somewhat vague. Clear examples of source-based terms are the Icelandic and Faroese words *appelsínugulur* “orange” and *brandgulur* “orange”, literally meaning “orange yellow”, where “orange” refers to the fruit rather than the color, and “fire yellow”, respectively.

Table 2

Numbers of shared etymologies within language families.

Language family												
Germanic (9)	9	8/1	9	5/4	9	9	9	7/1/1	3/3/2/1	4/2/1/1/1	7/2	
West Germanic (4)	4	3/1	4	4	4	4	4	4	2/1/1	3/1	2/2	
North Germanic (5)	5	5	5	5	5	5	5	3/1/1	3/2	2/1/1/1	5	
Italic (16)	14/2	15/1	8/4/2/2	10/4/2	16	11/4/1	11/2/1/1/1	14/1/1	9/3/2/1/1	16	15/1	
Western Romance (12)	11/1	11/1	5/3/2/2	7/4/1	12	8/4	8/2/1/1	11/1	6/3/2/1	12	12	
Central Romance (2)	2	2	2	2	2	2	2	2	2	2	2	
Southern Romance (1)	1	1	1	1	1	1	1	1	1	1	1	
Eastern Romance (1)	1	1	1	1	1	1	1	1	1	1	1	
Celtic (6)	3/3	6	3/2/1	3/3	3/3	3/3	3/2/1	4/1/1	4/1/1	2/2/1/1	5/1	
Hellenic (1)	1	1	1	1	1	1	1	1	1	1	1	
Albanian (1)	1	1	1	1	1	1	1	1	1	1	1	
Balto-Slavic (12)	12	10/1/1	8/1/1/1/1	11/1	12	5/4/1/1/1	3/2/2/2/1/1/1	9/1/1/1	6/5/1	12	4/4/2/2	
West Slavic (3)	3	3	3	3	3	2/1	2/1	2/1	2/1	3	2/1	
South Slavic (4)	4	4	3/1	3/1	4	2/2	2/1/1	2/1/1	2/1/1	4	4	
East Slavic (3)	3	3	2/1	3	3	3	3	3	3	3	3	
Baltic (2)	2	2	1/1	2	2	1/1	1/1	2	1/1	2	2	
Uralic (3)	2/1	2/1	2/1	2/1	1/1/1	2/1	2/1	3	2/1	2/1	1/1/1	
Finnic (2)	2	2	2	2	1/1	2	1/1	2	2	1/1	1/1	
Ugric (1)	1	1	1	1	1	1	1	1	1	1	1	
Vasconic (1)	1	1	1	1	1	1	1	1	1	1	1	
Afro-Asiatic (1)	1	1	1	1	1	1	1	1	1	1	1	
Semitic (1)	1	1	1	1	1	1	1	1	1	1	1	

As for the similarities between color terms within language families, Table 2 above gives an overview of the origins of the different color terms within certain language families. The number in brackets behind each family is the number of languages in the

sample within this family. The numbers in the cells indicate how many languages of a certain family share a color term with similar origins. For instance, the yellow column contains “5/4” for the Germanic family. This means that five of the nine Germanic color terms in the sample have the same etymology, while the remaining four also share an etymology. Of course, it would be more informative to perform statistical tests with these data, but as this paper is the first of its kind and its aim is largely explorative, statistics are beyond the scope of this paper.

As can be seen from the table, the homogeneity of color terms within language families varies. The words for the colors green, white and black quite often appear to have etymologies shared by all languages within a language family, whereas for other colors, such as brown, the opposite is true. For some languages, such as Greek and Albanian, this table is not very informative, of course, as they are not closely related to any other languages in the sample. For these languages, it is more informative to note that many of their color terms are etymologically unique within the sample and do not have cognates within other subfamilies. Additionally, Albanian *verdhë* “yellow” and *blerë* “green” are unique forms for the colors they refer to.

When looking at some of the language families in the sample, one can see that the languages within the family pattern together at an even lower level of classification that was not even included in the analysis. The subdivision of the Celtic languages into the Goidelic languages of Manx, Irish and Gaelic and the Brittonic languages of Welsh, Cornish and Breton, for instance, was not included in the analysis. One can see, however, that the words for the colors red, yellow, green, blue, white and brown of the Goidelic languages have the same etymological origins, as do the words for the colors yellow, green, blue, white and black of the Brittonic languages (Matasovic, 2008).

Although to a much lesser degree than its language family, the geographical location of a language can also determine the color terms it has. For instance, even though Frisian is much more closely related to English than it is to Dutch, its color terms are highly similar to those of Dutch, to the point where their forms for the color purple, which have the same etymology, are unique among the languages of Europe. Similarly, the Breton words for pink and gray are clearly borrowed from French due to their geographical proximity. Additionally, the Lithuanian, Latvian and Estonian words for yellow all originate in Proto-Balto-Slavic **gilʔtos* “species of moss” (Derksen, 2014).

Two languages with a high number of clearly borrowed color terms are Basque and Maltese. These two languages are completely unrelated to, but through geographical proximity heavily influenced by Spanish and Italian, respectively.

Interestingly, both languages have native color terms for the upper six colors of the color hierarchy and borrowed terms for the lower five, with the exception of Basque *berde* “green”, which is also borrowed. In Maltese, the borrowed color terms, which do not inflect, are in heavy contrast with the native color terms, which decline according to Arabic nonconcatenative morphology. For instance, *isfar* “yellow” becomes *saфра* when feminine and *sofor* when plural (Simpson, 2009).

After closely considering the forms and etymologies of color terms across Europe, one can also compare the color systems of languages as a whole and see if different languages in the sample distinguish different numbers of colors. When examining the sample, it appears that all languages distinguish the same eleven colors. One could argue, however, that Danish, Faroese, Manx and Irish, for instance, do not distinguish pink as a separate color due to their lack of a basic term for this color. Still, these lexical gaps appear to be very rare in the sample and do not hint at clear crosslinguistic patterns.

Technically, there are also languages within the sample that distinguish more than eleven colors. For instance, both Greek and Russian have separate terms for light blue, *galázio* and *golubój*, respectively, and Hungarian has a separate term for a darker shade of red, *vörös*. It appears that Russian speakers actually perceive light and dark blue as two different colors as a result of this (Winawer et al., 2007). Although some of these languages, especially Russian, can arguably be considered to distinguish more than eleven colors, this additional distinction does apparently not occur in related languages, such as the other Slavic languages besides Russian.

One case where a distinction of additional colors occurs in multiple languages across a language family concerns the languages Manx, Irish and Gaelic. For the color green, Manx has the word *geayney* and Irish and Gaelic have *uaine*, but these terms appear to be mostly reserved for dyed, artificially colored or light green objects. For naturally green things, such as grass or trees, Manx uses the word *glass* and Irish and Gaelic use *glas*, having the same etymology as the word for blue in the other Celtic languages. Although it remains disputable whether speakers of these languages perceive different shades of green as categorically different colors, this distinction forms an interesting pattern within the Celtic language family.

4. Discussion

In the previous, an analysis was provided for the origins of color terms across the languages of Europe, focusing on the etymologies of color terms themselves as well as a comparison of the color terms and color categories of languages within language

families. Centered around three corresponding research questions, this paper analyzed a sample of color terms gathered from fifty European languages. To a large extent, the hypotheses were confirmed. Most notably, for many of the languages in the sample, the terms for colors higher up in the color hierarchy were inherited from their respective proto-language, remained semantically stable throughout the years and were often shared between members of a language family.

As can be seen from the sample, languages can have several options to acquire color terms. They can inherit a term referring to the same color from their respective proto-language, such as Dutch *rood* “red”, inherit a term previously referring to a different color, such as Lithuanian *mėlyna* “blue”, morphologically derive a term from another term already present in the language, such as Danish *lyserød* “pink”, use a source-based term, such as Faroese *brandgulur* “orange” or, finally, borrow a term from a neighboring language, such as Basque *more* “purple”.

When looking at the origins of terms for the upper colors of the hierarchy, languages appear to mostly inherit them. Conversely, whenever languages employ one of the other strategies, it is usually in order to acquire a term for a color lower in the hierarchy. These tendencies appear to verify the suggested universality of the color hierarchy, as the semantic stability and inheritance from proto-languages seen with the terms for the upper colors of the hierarchy suggests that these colors have been consistently codified in a language throughout different stages of its development, which, according to the color hierarchy, is a prerequisite for the lower colors of the hierarchy to arise within a language.

These findings appear to be related to the findings that certain color terms, such as those for green, white and black, are very homogeneous within language families. For instance, the word for green has the same etymology for every single Italic language in the sample, because each of these languages acquired the term by inheriting it from their proto-language as a semantically stable word for green.

Conversely, the word for brown seems to have numerous different etymologies even among closely related languages, probably because languages employ a variety of strategies to acquire terms for this color, such as borrowing or using source-based terms. Of course, these observations do not apply to all language families in the sample, as among the Uralic languages, for instance, the word for green is not very homogeneous and, for example, the word for brown instead has the same etymology in all Germanic languages.

Generally, the hypothesis that more closely related languages will have similar color terms is also supported by the results. Although the homogeneity of color terms within language families varies greatly for different colors and from family to family,

it can be safely stated that the family a language belongs to, as was expected, plays a much bigger role in determining the color terms of this language than the geographical location of the language, as only a few languages in the sample display color terms that were borrowed from languages in the area.

As was also expected, all languages in the sample had a term of some sorts for the eleven colors taken into consideration here. Depending on one's definition, there might be a few languages that distinguish fewer colors, such as Danish and Manx, and there might be a few that distinguish more, such as Russian and Hungarian, but these lacking or additional distinctions appear to be an exception rather than a rule. As a result, there are, against expectations, no observable patterns of certain lacking or additional color distinctions shared between closely related languages, except maybe for the distinction of different hues of green in some Celtic languages.

One limitation to the current study concerns the sample size. Due to practical reasons, a limited selection of European languages had to be chosen, but there are of course many more distinct languages and dialects spoken all across Europe. Despite the limited sample size, however, the sample was compiled in such a way that it would proportionally represent all different language families and subfamilies spoken in Europe as well as cover the continent geographically, in order to maximize the validity of conclusions based on the sample.

Another limitation concerns the classification of the languages in the sample into subfamilies. European languages can generally be classified down to considerably low levels, with each level comprising several smaller and increasingly specific subbranches. The classification as it was used in this study was chosen for practical reasons and might be somewhat arbitrary in some ways. The effects of this on the results can be seen with, for instance, the Celtic languages, which share color terms down to a level that was not included in the analysis.

Sometimes, the retrieval of the appropriate color terms from the languages could also provide some difficulties. For instance, some languages in the sample had multiple words for certain colors. In these cases, the most basic or most frequently used term would be selected. However, since there were no concrete criteria, through which the appropriate term could be selected in these cases, this is arguably slightly problematic, especially considering that the selection of one term over another could greatly influence the proportions of shared etymologies shown in Table 2.

One might also argue that the approach that was taken in this study to search the literature for color terms is not completely sound. One could say, for instance, that it is not right to select the colors to search for in the literature beforehand, as this could bias the results towards more homogeneous findings. This concern would be very

valid if the sample consisted of highly diverse languages from all over the world with great differences in lexicon and, as a result, color terminology. As this sample consisted of European languages only, however, the expectation was that all languages in the sample would definitely have a term of some sort for the eleven colors searched for, as was stated before, which also proved to be the case.

This study is the first to directly relate color terminology to relationships between languages, so there is a lot of space for improvement and expansion in the form of follow-up studies. For instance, the numbers and proportions in Table 2 are to a certain degree informative, but it would be more interesting to see, as was stated before, if one could apply statistical analyses to such numbers, so that more definitive conclusions could be drawn from them.

Another possible follow-up study could be a comparative study of the processing of the different words for green by speakers of different Celtic languages to see whether they perceive these different shades of green as categorically different colors. This study could show whether or not discrepancies in the number of distinguished colors in a language can persist across members of the same language family.

5. Conclusion

In conclusion, this study attempted to combine several areas of research by relating previous crosslinguistic research on color terminology to language family classifications within historical linguistics. The study seems to offer additional evidence for the truthfulness and universality of the color hierarchy, as the upper colors of the hierarchy display semantically stable inheritance from proto-languages and more homogeneity across closely related languages, whereas the lower colors are often subject to borrowing from language to language and less homogeneity across languages.

Aside from this, the study hopefully also offers some promising results in general that are valuable to the research area dealing with color terminology and lays a foundation for follow-up research. By combining several different areas of research, this study hopes to contribute to research on color terminology and subsequently offer more insight into the subtle mechanics that aid humans in mapping the visual perception of color onto language.

References

- Abondolo, D. (2006a). Finnish. In *The Uralic Languages* (pp. 149-183). Abingdon-on-Thames: Routledge.
- Abondolo, D. (2006b). Hungarian. In *The Uralic Languages* (pp. 428-456). Abingdon-on-Thames: Routledge.
- Andersson, E. (2002). Swedish. In E. König & J. van der Auwera (Eds.), *The Germanic Languages* (pp. 271-312). Abingdon-on-Thames: Routledge.
- Askedal, J. O. (2002). Norwegian. In E. König & J. van der Auwera (Eds.), *The Germanic Languages* (pp. 219-270). Abingdon-on-Thames: Routledge.
- Balode, L., & Holvoet, A. (2001a). The Latvian language and its dialects. In Ö. Dahl & M. Koptjevskaja-Tamm (Eds.), *Circum-Baltic Languages* (pp. 3-40). Amsterdam: John Benjamins Publishing.
- Balode, L., & Holvoet, A. (2001b). The Lithuanian language and its dialects. In Ö. Dahl & M. Koptjevskaja-Tamm (Eds.), *Circum-Baltic Languages* (pp. 41-80). Amsterdam: John Benjamins Publishing.
- Barnes, M. P., & Weyhe, E. (2002). Faroese. In E. König & J. van der Auwera (Eds.), *The Germanic Languages* (pp. 190-218). Abingdon-on-Thames: Routledge.
- Benincà, P., Parry, M., & Pescarini, D. (2016). The dialects of northern Italy. In A. Ledgeway & M. Maiden (Eds.), *The Oxford guide to the Romance languages* (pp. 185-205). Oxford: Oxford University Press.
- Berlin, B., & Kay, P. (1991). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.
- Broderick, G. (2002). Manx. In M. Ball & J. Fife (Eds.), *The Celtic Languages* (pp. 228-286). Abingdon-on-Thames: Routledge.
- Brown, W. (2002). Serbo-Croat. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 306-387). Abingdon-on-Thames: Routledge.
- Buser, P., & Imbert, M. (1992). *Vision*. Cambridge: Bradford.
- Davies, I. R., & Corbett, G. G. (1997). A cross-cultural study of colour grouping: Evidence for weak linguistic relativity. *British journal of Psychology*, 88(3), 493-517.
- de Schutter, G. (2002). Dutch. In E. König & J. van der Auwera (Eds.), *The Germanic Languages* (pp. 439-477). Abingdon-on-Thames: Routledge.
- de Vaan, M. A. C. (2008). *Etymological dictionary of Latin and the other Italic languages*. Leiden: Brill.
- de Valk, J. M., Wnuk, E., Huisman, J. L., & Majid, A. (2017). Odor-color associations differ with verbal descriptors for odors: a comparison of three linguistically diverse groups. *Psychonomic bulletin & review*, 24(4), 1171-1179.

- Demiraj, S. (2015). Albanian. In A. Giacalone Ramat & P. Ramat (Eds.), *The Indo-European Languages* (pp. 480-501). Abingdon-on-Thames: Routledge.
- Derksen, R. (2007). *Etymological dictionary of the Slavic inherited lexicon*. Leiden: Brill.
- Derksen, R. (2014). *Etymological dictionary of the Baltic inherited lexicon*. Leiden: Brill.
- Eisenberg, P. (2002). German. In E. König & J. van der Auwera (Eds.), *The Germanic Languages* (pp. 349-387). Abingdon-on-Thames: Routledge.
- Epps, P., & Huehnergard, J. (2013). Introduction: Contact among genetically related languages. *Journal of language contact*, 6(2), 209-219.
- Friedman, V. A. (2002). Macedonian. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 249-305). Abingdon-on-Thames: Routledge.
- George, K. (2002). Cornish. In M. Ball & J. Fife (Eds.), *The Celtic Languages* (pp. 410-468). Abingdon-on-Thames: Routledge.
- Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., Gibson, M., Piantadosi, S. T., & Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, 114(40), 10785-10790.
- Gillies, W. (2002). Scottish Gaelic. In M. Ball & J. Fife (Eds.), *The Celtic Languages* (pp. 145-227). Abingdon-on-Thames: Routledge.
- Haberland, H. (2002). Danish. In E. König & J. van der Auwera (Eds.), *The Germanic Languages* (pp. 313-348). Abingdon-on-Thames: Routledge.
- Haiman, J. (1988). Rhaeto-Romance. In M. Harris & N. Vincent (Eds.), *The Romance Languages* (pp. 351-390). Abingdon-on-Thames: Routledge.
- Harris, M. (1988). French. In M. Harris & N. Vincent (Eds.), *The Romance Languages* (pp. 209-245). Abingdon-on-Thames: Routledge.
- Hoekstra, J., & Tiersma, P. M. (2002). Frisian. In E. König & J. van der Auwera (Eds.), *The Germanic Languages* (pp. 505-531). Abingdon-on-Thames: Routledge.
- Hoenigswald, H. M. (2015). Greek. In A. Giacalone Ramat & P. Ramat (Eds.), *The Indo-European Languages* (pp. 228-260). Abingdon-on-Thames: Routledge.
- Jones, M. (1988). Sardinian. In M. Harris & N. Vincent (Eds.), *The Romance Languages* (pp. 314-350). Abingdon-on-Thames: Routledge.
- Kay, P., Berlin, B., Maffi, L., & Merrifield, W. R. (1997). Color naming across languages. *Color categories in thought and language*, 21, 58.
- Kay, P., Berlin, B., Maffi, L., Merrifield, W. R., & Cook, R. (2009). *The world color survey*. Stanford: CSLI Publications.
- Kaye, A. S., & Rosenhouse, J. (1997). Arabic Dialects and Maltese. In R. Hetzron (Ed.), *The Semitic Languages* (pp. 263-311). Abingdon-on-Thames: Routledge.

- König, E. (2002). English. In E. König & J. van der Auwera (Eds.), *The Germanic Languages* (pp. 532-565). Abingdon-on-Thames: Routledge.
- Kroonen, G. (2013). *Etymological dictionary of Proto-Germanic*. Leiden: Brill.
- Ledgeway, A. (2016). The dialects of southern Italy. In A. Ledgeway & M. Maiden (Eds.), *The Oxford guide to the Romance languages* (pp. 246-269). Oxford: Oxford University Press.
- Livingstone, M. (2002). *Vision and Art: The Biology of Seeing*. New York: Abrams.
- Mac Eoin, G. (2002). Irish. In M. Ball & J. Fife (Eds.), *The Celtic Languages* (pp. 101-144). Abingdon-on-Thames: Routledge.
- Mallinson, G. (1988). Rumanian. In M. Harris & N. Vincent (Eds.), *The Romance Languages* (pp. 391-419). Abingdon-on-Thames: Routledge.
- Matasovic, R. (2008). *Etymological dictionary of Proto-Celtic*. Leiden: Brill.
- Mayo, P. (2002). Belorussian. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 887-946). Abingdon-on-Thames: Routledge.
- Parkinson, S. (1988). Portuguese. In M. Harris & N. Vincent (Eds.), *The Romance Languages* (pp. 131-169). Abingdon-on-Thames: Routledge.
- Priestly, T. M. S. (2002). Slovene. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 388-454). Abingdon-on-Thames: Routledge.
- Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences*, 13(10), 439-446.
- Regier, T., Kay, P., & Khetarpal, N. (2009). Color naming and the shape of color space. *Language*, 85(4), 884-892.
- Renfrew, C. (1989). The origins of Indo-European languages. *Scientific American*, 261(4), 106-115.
- Rothstein, R. A. (2002). Polish. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 686-758). Abingdon-on-Thames: Routledge.
- Scatton, E. A. (2002). Bulgarian. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 188-248). Abingdon-on-Thames: Routledge.
- Shevelov, G. Y. (2002). Ukrainian. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 947-998). Abingdon-on-Thames: Routledge.
- Short, D. (2002a). Czech. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 455-532). Abingdon-on-Thames: Routledge.
- Short, D. (2002b). Slovak. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 533-592). Abingdon-on-Thames: Routledge.
- Simpson, A. K. (2009). *The origin and development of nonconcatenative morphology* (Doctoral dissertation). University of California.

- Stephens, J. (2002). Breton. In M. Ball & J. Fife (Eds.), *The Celtic Languages* (pp. 349-409). Abingdon-on-Thames: Routledge.
- Thráinsson, H. (2002). Icelandic. In E. König & J. van der Auwera (Eds.), *The Germanic Languages* (pp. 142-189). Abingdon-on-Thames: Routledge.
- Timberlake, A. (2002). Russian. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 827-886). Abingdon-on-Thames: Routledge.
- Tuten, D. N., Pato, E., & Schwarzwald, O. R. (2016). Spanish, Astur-Leonese, Navarro-Aragonese, Judaeo-Spanish. In A. Ledgeway & M. Maiden (Eds.), *The Oxford guide to the Romance languages* (pp. 382-410). Oxford: Oxford University Press.
- Viitso, T. R. (2006). Estonian. In D. Abondolo (Ed.), *The Uralic Languages* (pp. 115-148). Abingdon-on-Thames: Routledge.
- Vincent, N. (1988). Italian. In M. Harris & N. Vincent (Eds.), *The Romance Languages* (pp. 279-313). Abingdon-on-Thames: Routledge.
- Watkins, T. A. (2002). Welsh. In M. Ball & J. Fife (Eds.), *The Celtic Languages* (pp. 289-348). Abingdon-on-Thames: Routledge.
- Wheeler, M. W. (1988a). Catalan. In M. Harris & N. Vincent (Eds.), *The Romance Languages* (pp. 170-208). Abingdon-on-Thames: Routledge.
- Wheeler, M. W. (1988b). Occitan. In M. Harris & N. Vincent (Eds.), *The Romance Languages* (pp. 246-278). Abingdon-on-Thames: Routledge.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, 104(19), 7780-7785.
- Zuazo, K. (2019). *Standard Basque and Its Dialects*. Abingdon-on-Thames: Routledge.

I have written this paper for the course Linguistic Universals and Diversity from the master General Linguistics at Radboud University that I followed in 2019. The original paper contained more elaborate results, which had to be left out due to length issues. If interested in these additional results, please contact the author.







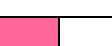



Appendix

Sample of color terms and references

Table A1 contains the sample of color terms, whereas Table A2 contains the used sources. Note that the references for these sources are to be found in the reference list above.

Table A1

Sample of color terms of the fifty languages.

Language											
English	red	orange	yellow	green	blue	purple	pink	white	grey	black	brown
Frisian	read	oranje	giel	grien	blau	pears	rôze	wyt	griis	swart	brún
Dutch	rood	oranje	geel	groen	blauw	paars	roze	wit	grijs	zwart	bruin
German	rot	orange	gelb	grün	blau	violett	rosa	weiß	grau	schwarz	braun
Danish	rød	orange	gul	grøn	blå	lilla	lyserød	hvid	grå	sort	brun
Swedish	röd	orange	gul	grön	blå	lila	skär	vit	grå	svart	brun
Norwegian	rød	oransje	gul	grønn	blå	lilla	rosa	hvit	grå	svart	brun
Icelandic	rauður	appelsínugulur	gulur	grænn	blár	purpuralitur	bleikur	hvítur	grár	svartur	brúnn
Faroese	reyður	brandgulur	gulur	grønur	bláur	korkalitur	ljósareyður	hvitur	gráur	svartur	brúnur
Romansh	cotschen	oranscha	mellen	verd	blau	violet	rosa	alv	grisch	nair	brin
Ladin	cueciun	orancen	śal	vërt	brum	polpra	rosa	blanch	grisc	fosch	ros
Friulian	ros	narancin	zâl	vert	blâv	viole	rose	blanc	gris	neri	moron
Lombard	ros	aranciú	zald	verd	blö	viola	roza	bianch	gris	négher	marú
Piedmontese	ross	arancion	giàun	verd	bleu	viola	reusa	bianch	gris	nèir	maròn
French	rouge	orange	jaune	vert	bleu	violet	rose	blanc	gris	noir	marron
Occitan	roge	irange	jaune	verd	blu	violet	ròse	blanc	gris	negre	brun
Catalan	roig	taronja	groc	verd	blau	porpra	rosa	blanc	gris	negre	marró
Aragonese	royo	narancha	amariello	berde	azul	purpurencó	rosa	blanco	griso	negro	pardo
Asturian	bermeyu	naránxa	mariellu	verde	azul	moráu	rosa	blancu	gris	negru	marrón
Spanish	rojo	naranja	amarillo	verde	azul	morado	rosa	blanco	gris	negro	marrón
Portuguese	vermelho	laranja	amarelo	verde	azul	roxo	rosa	branco	gris	negro	marron
Italian	rosso	arancione	giallo	verde	blu	viola	rosa	bianco	grigio	nero	marrone
Sicilian	russu	arancia	giarnu	viridi	bru	viola	rosa	biancu	griciu	níuru	marruni
Sardinian	ruju	aranzu	groggu	birde	blu	viola	rosa	biancu	murru	nigheddu	tabachinu
Romanian	roșu	portocaliu	galben	verde	albastru	mov	roz	alb	gri	negru	maro
Manx	jiarg	jiarg-bwee	bwee	geayne	gorrym	jiarg-gorrym	jiarg-bane	bane	lheeah	doo	dhoan
Irish	dearg	oráiste	buí	uaine	gorm	corcra	bándearg	bán	liath	dubh	donn
Gaelic	dearg	orains	buidhe	uaine	gorm	purpaidh	pinc	bàn	liath	dubh	donn
Welsh	coch	oren	melyn	gwyRDD	glas	porffor	pinc	gwyn	llwyd	du	gwrn
Cornish	rudh	rudhvelyn	melyn	gwyrdh	glas	purpur	kigliw	gwynn	loos	du	gorm
Breton	ruz	orañjez	melen	gwer	glas	limestra	roz	gwenn	gris	du	melegen
Greek	kókkino	portokali	kitrino	prásino	ble	mov	roz	áspro	gkri	mávro	kafé
Albanian	kuq	portokalltë	verdhë	blerë	kaltër	purpurtë	pembë	bardhë	hirtë	zi	kaftë
Polish	czerwony	pomarańczowy	żółty	zielony	niebieski	purpurowy	różowy	biały	szary	czarny	brązowy
Czech	červená	oranžová	žlutá	zelená	modrá	fialová	růžová	bílá	šedá	černá	hnědá
Slovak	červená	oranžová	žltá	zelená	modrá	fialová	ružová	biela	šedá	čierna	hnedá
Slovene	rdeča	oranžna	rumena	zelena	modra	vijolična	roza	bela	siva	črna	rjava
Serbian	crven	narandžast	žut	zelen	modar	ljubičast	ružičast	beo	siv	crn	braon
Bulgarian	červén	oránžev	žlt	zelen	sin	violétov	rózov	bjal	siv	čeren	kafjáv
Macedonian	crven	portokalov	žolt	zelen	sin	purpuren	rozov	bel	siv	crn	kafen
Russian	krásnyj	oránževyj	žóltij	zeljónyj	sinij	púrpurnyj	rózovyj	bélyj	séryj	čórnyj	koríčnevij
Ukrainian	červónyj	oránževyj	žóvtij	zelényi	sýnij	púrpurovyj	roževyj	bílyj	síryj	čórnyj	korýčnevij
Belarusian	čyrvóny	aránžavy	žóuty	zjaljóny	síni	purpúrny	ružóvy	bjély	šéry	čórny	karýčnjevy
Lithuanian	raudona	oranžinė	geltona	žalia	mėlyna	violetinė	rožinė	balta	pilka	juoda	ruda
Latvian	sarkans	oranžs	dzeltens	zaļš	zils	purpursarkans	rozā	balts	pelēks	melns	brūns
Estonian	punane	oranž	kollane	roheline	sinine	purpurne	roosa	valge	hall	must	pruun
Finnish	punainen	oranssi	keltainen	vihreä	sininen	purppura	pinkki	valkea	harmaa	musta	ruskea

Hungarian	piros	narancsszín	sárga	zöld	kék	bíbor	rózsaszín	fehér	szürke	fekete	barna
Basque	gorri	laranja	hori	berde	urdin	more	arrosa	zuri	gris	beltz	marroi
Maltese	ahmar	orangjo	isfar	ahdar	ikhal	vjola	roża	abjad	griz	iswed	kannella

Table A2

References used for each of the fifty languages.

Language	Reference	Language	Reference
English	König, 2002	Manx	Broderick, 2002
Frisian	Hoekstra & Tiersma, 2002	Irish	Mac Eoin, 2002
Dutch	de Schutter, 2002	Gaelic	Gillies, 2002
German	Eisenberg, 2002	Welsh	Watkins, 2002
Danish	Haberland, 2002	Cornish	George, 2002
Swedish	Andersson, 2002	Breton	Stephens, 2002
Norwegian	Askedal, 2002	Greek	Hoenigswald, 2015
Icelandic	Thráinsson, 2002	Albanian	Demiraj, 2015
Faroese	Barnes & Weyhe, 2002	Polish	Rothstein, 2002
Romansh		Czech	Short, 2002a
Ladin	Haiman, 1988	Slovak	Short, 2002b
Friulian		Slovene	Priestly, 2002
Lombard		Serbian	Brown, 2002
Piedmontese	Benincà, Parry, & Pescarini, 2016	Bulgarian	Scatton, 2002
French	Harris, 1988	Macedonian	Friedman, 2002
Occitan	Wheeler, 1988b	Russian	Timberlake, 2002
Catalan	Wheeler, 1988a	Ukrainian	Shevelov, 2002
Aragonese		Belarusian	Mayo, 2002
Asturian	Tuten, Pato, & Schwarzwald, 2016	Lithuanian	Balode & Holvoet, 2001b
Spanish		Latvian	Balode & Holvoet, 2001a
Portuguese	Parkinson, 1988	Estonian	Viitso, 2006
Italian	Vincent, 1988	Finnish	Abondolo, 2006a
Sicilian	Ledgeway, 2016	Hungarian	Abondolo, 2006b
Sardinian	Jones, 1988	Basque	Zuazo, 2019
Romanian	Mallinson, 1988	Maltese	Kaye & Rosenhouse, 1997

Polari

Language as a means of concealment

R. (Roel) Stroo

A.A.K. (Bram) Peute

Master *General Linguistics*, Radboud University Nijmegen

Abstract: Until the passing of the Sexual Offences Act in 1967, homosexuality was a crime in the United Kingdom. In order to be able to talk freely about their sexuality, gay men spoke Polari, a language that has much in common with English but has a lexicon that is centered around the culture of the gay community. Its purpose was to include its speakers in the gay community, while people who did not speak it were kept out. In this paper, we investigated whether Polari was successful in this endeavor. We showed that Polari likely fulfilled its purpose of concealing the sexual identity of its speech community, as Polari does not sound very exotic, while the nature of the matters discussed remain hidden. We conducted our research by employing the method of distributing questionnaires that exposed the participants to a short conversation in Polari. We subsequently analyzed their perception of Polari through the answers they provided. It can be concluded from the results that Polari enabled its speakers to discuss topics which were not only socially deemed inappropriate but illegal as well, while still blending in with other English dialect speakers.

Key words: Polari, language and homosexuality, concealing identity, anti-language

1. Introduction

“So bona to vada... oh you! Your lovely eek and your lovely riah...” is one of the lines in Morrissey’s song “Piccadilly Palare”. These words belong to Polari, a secret language spoken among homosexual men in the London area in order to conceal their identity, as homosexuality used to be a crime in the United Kingdom until the passing of the Sexual Offences Act in 1967 (Higgins, 1996, p. 115). Since then, the status of Polari has changed: gay men no longer had to hide their sexual preferences, which led to the decline of the usage of Polari among homosexuals. At the same time, the Polari lexicon became public knowledge after its original purpose of hiding a person’s homosexuality had become obsolete. Nowadays, Polari is not being taught to new members of the gay community anymore and is said to be “moribund”, despite several

attempts by the gay community to revive the language (Taylor, 2007, p. 32). As Polari was a language designed to conceal the identities of its speech community, it should be inconspicuous enough to not draw attention to its speakers but incomprehensible enough to not expose its speakers. These observations form the basis for the following research question: how did Polari fulfil its purpose of concealing the sexual identity of its speech community? And if it is the case that people recognize or understand Polari, are there any noticeable patterns in terms of age, gender, or sexuality?

Polari can be considered an example of an *anti-language*: a language spoken by an *anti-society* (Halliday, 1976), an alternative society within a society, with its own social structure and its own lexicon. These anti-societies, designed to resist the mainstream culture, are usually built around illegal activities such as drug dealing, stealing, squatting, and, in this case, acts of homosexuality. Logically, this means that anti-languages are centered around the activities of the anti-society that they are spoken by: anti-languages tend to have the same grammar as mainstream languages, but have lexicons based on their activities. For example, the Polari lexicon is characterized by varieties of terms for sex, body parts, and people (Taylor, 2007, p. 20).

Polari is not the only language that serves to distinguish groups of different sexual orientations. In fact, language and sexual orientation have always been connected. Another example of this connection is the isiNgcqumo language, spoken by gay people in the ethnic Zulu community in South Africa (Rudwick, 2010). IsiNgcqumo shows that identification through language is not exclusively used to distinguish people with the same sexual orientation, but also to identify as a member of a certain ethnographic group. It is worth noting that literature on this subject is scarce, mostly due to the fact that the number of Polari speakers has rapidly declined over the years as it does not serve its purpose anymore and, as a result, empirical research on Polari has become increasingly difficult, if not impossible.

2. Method

In order to investigate the current status of Polari, we designed a questionnaire (included in the appendix) that consisted of several questions regarding a sound fragment from the short film *“Putting on the dish”* (Fairbairn & Eccleston, 2015), in which two gay men have a conversation in Polari.

2.1 Participants

The questionnaire was spread among people from the UK through Facebook, but the gross of participants came from the forum casualUK, a forum hosted on the social media website Reddit. The 131 participants were informed to only reply if they were from the UK; this was considered a vital condition, as we were looking for native evaluations of a local dialect. The participants were also asked to fill out their age, gender, sexual orientation, and hometown, along with whether they considered their hometown to be a village, a town, or a city. This was done to be able to see if these demographic differences influenced participants' knowledge of Polari. The descriptive information about the participants is included below.

Table 1

Descriptive participant information.

Variable	N	%
Age		
20-	21	16,0
21-30	58	44,3
31-40	35	26,7
41-50	10	7,6
50+	7	5,3
Total	131	≈100%
Gender		
Male	79	60,3
Female	49	37,4
Other	3	2,3
Total	131	100%
Sexual orientation		
Straight	96	73,3
Gay	5	3,8
Bisexual	15	11,5
Other	15	11,5
Total	131	≈100%
Size of hometown		
City	45	34,4%
Town	62	47,3%
Village	24	18,3%
Total	131	100%

2.2 Materials

As mentioned above, the sound fragment to which the participants were exposed, was taken from the short film *“Putting on the dish”* (Fairbairn & Eccleston, 2015), which takes place in 1962 and is more or less dedicated to the usage of Polari in that era. We chose this sound fragment as the dialogue in it is neither too obvious nor too obscure, and due to the practical reason that the fragment was readily available, which is not evident for a secret language that has been declining since the 1970s. There was no video material involved and the participants could listen to the fragment at home on their own devices; there were no advanced sound systems of any kind required.

2.3 Procedure

After filling out a questionnaire about their personal information, the participants were exposed to the Polari sound fragment. Then, the participants rated their own comprehension of the dialogue on a scale from 1-10, with 1 being completely incomprehensible and 10 being completely comprehensible, and listed the words that came to mind while listening to the fragment in a think-aloud task. Finally, the participants were asked what they thought the dialogue was about and whether they recognized the language variety.

2.4 Analysis

The 131 participants were first organized in three groups: people who recognized Polari and identified it as such, people who thought they recognized it but misidentified the language variety, and people who did not recognize it. The results of the think-aloud test were then qualitatively analyzed per group. Several quantitative analyses were conducted as well. A one-way ANOVA was used to see if there was a significant effect of age on the recognition of Polari. A second one-way ANOVA was used to see if there was a significant effect of the self-evaluated comprehensibility scores on the recognition of Polari. The effects of sexuality, gender, and town size on the recognition of Polari were all analyzed with likelihood ratios.

3. Results

3.1 Qualitative analysis

The group of participants ($n = 20$, 15.3%) that understood the dialogue and recognized it as Polari agreed more or less on the content of the dialogue, while also giving similar responses to the think-aloud part: many participants associated the fragment with the Cockney dialect, sexual acts, and popular comedian Kenneth Williams, who played an important role in the integration of Polari in the UK after the decriminalization of homosexuality in 1967 (Higgins, 1967, p. 86). Most participants in this group were able to (fairly) accurately describe the dialogue, even the heterosexual participants.

More interesting is the large group of participants who reported that they understood the dialogue ($n = 58$, 44.3%), but could not tell that Polari was the language variety in question. Some participants were actually able to accurately retell the story, however, they often reported the dialogue being about a woman who was detained for performing a sexual act, while the characters are actually talking about a man; this is because Polari uses female pronouns to refer to men. These participants usually rated their own comprehension highly and gave extensive descriptions of the dialogue, however, they still missed the important detail that the dialogue was about a man, which shows that Polari still has the ability to conceal certain aspects of the topic discussed. In the think-aloud part, the participants in this group frequently associated the fragment with sex and the Cockney dialect, which is not surprising considering that Polari is derived from this dialect and the lexicon is centered around sexual activity, which was also a topic of conversation in the dialogue.

The last group consisted of participants that answered that they did not understand the conversation at all ($n = 53$, 40.5%). When asked about the content of the dialogue, many of the participants reported that the dialogue was about a woman and that it involved sexual acts, but could not put the pieces of the story together. As for the think-aloud part, this group provided a large variety of answers, but very few of them related to Polari itself. Most of these answers are rather broad and relate to everyday concepts, such as pub, banter and casual conversations.

3.2 Quantitative analysis

From a one-way ANOVA it was shown that there is a significant effect of age on correctly identifying the language fragment as Polari ($F(2, 43) = 7.37$, $p = .002$, $\eta^2 = .13$). The average age of participants who identified Polari correctly was 39.30 ($SD = 13.96$),

the average age of people who misidentified Polari was 29.19 ($SD = 8.37$), and the average age of people who did not recognize it was 28.38 ($SD = 9.32$).

On grounds of Hochberg's GT2 it can be concluded that participants who recognized Polari were significantly older than participants who misidentified it ($p = .00$), as well as participants who did not recognize it ($p = 0.00$). There was no significant age difference between participants who misidentified it and those who did not recognize it ($p = .96$).

From a One-way ANOVA it was shown that there was no significant effect of the self-evaluated comprehensibility score on correctly identifying the language fragment as Polari ($F(2, 128), p = .098, \eta^2 = .04$). The average comprehensibility score of participants who identified Polari correctly was 5.30 ($SD = 2.23$), the average comprehensibility score of people who misidentified Polari was 5.50 ($SD = 2.31$), and the average comprehensibility score of people who did not recognize it was 5.64 ($SD = 1.83$).

On grounds of Hochberg's GT2 it can be concluded that participants who recognized Polari had on average no significantly higher comprehensibility scores than participants who misidentified it ($p = .98$). This was also found when compared to participants who did not recognize Polari ($p = .56$). There was no significant difference between the comprehensibility scores of participants who misidentified and those who did not recognize the language variety ($p = .10$) either.

The second one-way ANOVA showed that the demographic distinctions made in the data had no effect on whether participants recognized the language fragment as Polari. There is no significant relation between the sexuality of the participants and the recognition of Polari ($\chi^2(10, n = 131) = 11.80, p > 0.05$). There is also no significant relation between the gender of participants and the recognition of Polari ($\chi^2(4, n = 131) = 6.01, p > .05$), and, finally, there was no significant relation found between the size of the participants' hometown and the recognition of Polari ($\chi^2(4, n = 131) = 2.97, p > .05$).

4. Conclusion

Twenty out of 131 participants recognized and correctly identified Polari. They are significantly older than the participants who misidentified or did not recognize the language variety. Several participants pointed out that they know it from Kenneth Williams, who hosted the radio show *Round the Horne* with Hugh Paddick, which aired in the 1960s, where they would play the Polari-speaking characters Julian and Sandy. This might explain the effect of age on recognizing Polari, as older people have probably listened to this radio show. Interestingly, the self-evaluated

comprehensibility scores of this group did not significantly differ from people who misidentified or did not recognize Polari.

There were 8 out of 131 participants who claimed they recognized the language variety but misidentified it. They often compared it to the Cockney dialect, which is related to Polari, and could tell what large parts of the dialogue were about. However, they were still fooled by the pronoun use of Polari, as they all thought the dialogue was about a woman.

The remaining participants, 53 out of 131, did not recognize the language variety. They could sometimes tell it was about a woman and sexual activities, but that was the extent of their comprehension. They associated the language variety with pub talk, banter, and other everyday concepts.

This clearly illustrates the concealing power of Polari: Polari is often seen as a related language variety, which makes its users blend in with other English dialect speakers, while it is still able to conceal the identity of its speakers from people who think they understand it. When homosexuality became legal, Polari went extinct and has no speakers left. Nowadays, knowing about Polari is not tied to sexuality, gender, or the size of one's hometown, but rather to having listened to an old radio show. In the heyday of Polari it probably fulfilled its role perfectly: it allowed its speakers to inconspicuously discuss their sexual exploits, deemed illegal by close-minded rulers.

References

- Fairbairn, B. & Eccleston, K. [Brian and Karl]. (2015, June 24). *Putting on the Dish: How gay men used to speak - A short film in Polari*. [Video file]. Retrieved from: <https://www.youtube.com/watch?v=Y8yEH8TZUsk>.
- Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. London: Edward Arnold.
- Higgins, P. (1996). *Heterosexual Dictatorship: Male Homosexuality in Postwar Britain*. London: Fourth Estate.
- Rudwick, S. (2010). 'Gay and Zulu, we speak isiNgqumo': Ethnolinguistic identity constructions. *Transformation: Critical Perspectives on Southern Africa*, 74(1), 112-134. <https://doi.org/10.1353/trn.2010.0016>.
- Taylor, H. (2007). *Polari: A sociohistorical study of the life and decline of a secret language*. Bachelor Thesis, University of Manchester.
- Parliament, U. K. (1967). Sexual offences act 1967. doi: https://www.legislation.gov.uk/ukpga/1967/60/pdfs/ukpga_19670060_en.pdf.

Appendix

Questionnaire

Section 1

Language Evaluation Task

!PLEASE ONLY REPLY IF YOU ARE FROM THE UK!

You need to listen to a soundfile, so please make sure you are in an environment where this is possible

If you feel uncomfortable answering a question, please fill in a backslash \

Section 2

About you

What is your age?

...

What gender do you identify with most?

- Male
- Female
- Other

What is your sexual orientation?

...

What is your hometown generally considered to be?

- Village
- Town
- City

What is the name of your hometown?

...

Section 3

Sound fragment

This is the important part

Please listen to the following sound fragment, and write down all the words that come to mind while listening to it

-Transcript of sound file-

- > I was seeing this HP from Sheffield once. Plates the size of bowling pins, I thought I was in for a real bona charvering.
- < Nada to varda in the larder?
- > Oh, bijou. 'You needn't put the brandy on for that,' I said when I saw it. Mind you, she was heavy on the letch water. I had to use the Daz to get her Maria out my libbage.
- < Oh, vile. Has she always been that way then, Phyllis?
- > She's a walking meat rack. Real fantabulosa bit of hard. We used to act dicky together at the croaker's chovey. Noshed me off once while I was giving a fungus his drabs.
- < That's skill, that.
- > Oh she used to do it all the time. When we were at the exchange together she'd one lill on my colin and the other on the switch. She didn't even get off the palare pipe. Sad to think of her in the queer ken really.
- < What do you mean?
- > Well she'd a run in with the lily law, didn't she?
- < Oh dear.
- > Sharpie flashed his cartso in the carsey.
- < I hope she kept her ogles front.
- > Well she's got amblyopia, hasn't she? She can practically only vada sideways.
- < What did the beak say?
- > He was ever so harsh. Asked if she was sorry.
- < Was she?
- > Only that it wasn't worth the look she got.

Section 4

Final questions

How much of the dialogue were you able to comprehend?

	1	2	3	4	5	6	7	8	9	10	
Completely incomprehensible	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Completely comprehensible

What do you think the dialogue was about?

...

Did you recognize the language variety?

- Yes
- No

If you recognized it, what do you think it was?

...

Thank you for filling in this questionnaire!

Waarom woorden in klassen passen:

Semantiek in het nominale classificatiesysteem van het Avatime

Semantics in the noun class system of Avatime

S.E.M. (Sam) Theunissen

Bachelor *Taalwetenschap*, Radboud University Nijmegen

Abstract: Voor dit onderzoek is de semantiek van het classificatiesysteem in het Avatime onderzocht, een Kwa-taal die wordt gesproken in Ghana. In het Avatime worden naamwoorden, evenals in veel andere talen, onderverdeeld in een aantal klassen. Resultaten van eerdere onderzoeken naar andere talen laten zien dat deze nominale classificatiesystemen niet geheel arbitrair zijn, maar een grammaticale weergave van de cognitieve categorieën die sprekers op basis van hun ervaringen maken (Sagna, 2012). Over de semantiek achter de naamwoordsklassen van het Avatime is nog vrij weinig bekend, behalve dat de eerste klasse is bestemd voor woorden voor menselijke referenten en dat de zevende naamwoordsklasse enkel stofnamen bevat (Schuh, 1995; van Putten, 2014). In dit onderzoek ligt de focus op de rollen die de levendheid van referenten speelt in het classificatiesysteem en de intuïties die Avatime sprekers hierover hebben. Het onderzoek heeft plaatsgevonden in Vane, Ghana, waar veertig participanten hebben deelgenomen aan twee experimenten om de volgende onderzoeksvraag te beantwoorden: in hoeverre speelt de levendheid van referenten een rol in de actieve classificatie van naamwoorden door sprekers van het Avatime? In Experiment 1 moesten deelnemers uit zes afbeeldingen de referent kiezen die volgens hen het beste paste bij het pseudowoord dat ze zojuist hadden gehoord. In Experiment 2 werd participanten gevraagd uit twee pseudowoorden, de een uit de eerste en de ander uit de vierde naamwoordsklasse, het woord te kiezen dat volgens hen het beste paste bij de afbeelding die ze zagen. De resultaten van beide experimenten bleken niet significant te zijn. Hieruit kan worden geconcludeerd dat het semantische kenmerk van levendheid geen grote rol speelt in het taalgevoel dat Avatime sprekers hebben over de nominale classificatie van hun taal.

Trefwoorden: levendheid, Avatime, taal en cognitie, naamwoordsklassen, semantiek

1. Inleiding

1.1 *Avatime*

In dit artikel zal een scriptie-onderzoek worden beschreven dat is uitgevoerd bij sprekers van het Avatime, een taal die wordt gesproken in de Volta Region, in het zuidoosten van Ghana. Het Avatime is onderdeel van de Niger-Congo talenfamilie en wordt geclassificeerd als een Kwa-taal in de Ka-Togo tak (Hammarström, Forkel & Haspelmath, 2019). Het is niet geheel duidelijk door hoeveel personen het Avatime wordt gesproken, de schattingen lopen van 11.600 (Blench, 2006) tot 15.000 (van Putten, 2014) en 27.200 (Lewis, Simons, & Fennig, 2013). De Ghanese bevolking door wie het Avatime wordt gesproken, is verspreid over acht dorpen die samen de *Avatime traditional area* vormen. Een van deze dorpen is Vane, dit is de plaats waar het huidige onderzoek is voltrokken.

Het huidige onderzoek heeft betrekking op de naamwoordsklassen van het Avatime. De hoeveelheid klassen en de manier waarop deze in de grammatica van een taal worden uitgedrukt zijn taalafhankelijk. In het Avatime zijn het er, volgens Schuh (1995), dertien. Deze kunnen worden gegroepeerd tot zeven genders, waarvan er bij zes een onderscheid in getal wordt gemaakt – het zevende gender is bestemd voor *mass nouns*, of stofnamen, en maakt daarom geen onderscheid tussen enkelvoudige en meervoudige referenten (Schuh, 1995; van Putten, 2014).

Er zijn verschillende zinslementen waarin de klasse van een nomen tot uiting kan worden gebracht. In het Avatime gebeurt dat in de vorm van een prefix op het zelfstandig naamwoord en affixen van congruentie op andere elementen in de zin. Er zijn drie soorten elementen binnen een nominale constituent die overeenstemming vertonen in naamwoordsklasse. Dat zijn (in)definiete lidwoorden, demonstratieven en telwoorden (van Putten, 2014). Daarnaast is er ook congruentie met onafhankelijke persoonlijke pronomina en subjectmarkeerders. Welke vormen de congruerende zinsdelen aan kunnen nemen, is te zien in Tabel 1.

Tabel 1

Naamwoordsklassen en congruerende elementen. Hoofdletters staan voor klanken die afhankelijk zijn van klinkerharmonie. Herdrukt van “Les 5: nominale constituenten” [Ongepubliceerde handout] van S. van Putten, 2015.

<i>Klasse</i>	<i>Prefix</i>	<i>Definiet lidwoord</i>	<i>Demonstratief</i>	<i>Indefiniet lidwoord</i>	<i>Telwoord</i>	<i>Persoonlijk pronomen</i>
1s	O-/Ø	-(y)E	líyè/lé-lò	ɔ-tɔ	to-le ‘één’	yɛ
1p	bA-/Ø	-a	bá-yà/ bá-lò	a-tɔ	tja-ba ‘twee’	ba
2s	Ò-	-LO	ló-	ɔ-tɔ	to-	lɔ
2p	Ì-	-LE	lé-	j/ɛ-tɔ	tɪ-	lɛ
3s	ɪ-	-LE	lé-	ɛ-tɔ	ti-	lɛ
3p	A-	-La	lá-	a-tɔ	tA-	la
4s	kɪ-	-(y)E	ké-	j/ɛ-tɔ	ti-	kɛ
4p	bɪ-	-E	bé-	j/ɛ-tɔ	tU(I)-	bɛ
5s	kU-	-O	kó-	ɔ-tɔ	tu-	kɔ
5p	bÀ-	-a	bá-	a-tɔ	tɪa-	ba
6s	kA-	-a	ká-	a-tɔ	ti-	ka
6p	kÛ-	-O	kó-	ɔ-tɔ	tU-	kɔ
7	sɪ-	-sE	sé-	ɛ-tɔ	ti-	sɛ

De hoofdletters in Tabel 1 verwijzen naar de klinkerharmonie zoals die van toepassing is in de fonologie van het Avatime, op basis van *Advanced Tongue Root* (ATR). In de context van het Avatime houdt deze klinkerharmonie in, dat de positie van de tong bij de uitspraak van klinkers in het pre- en het suffix afhankelijk is van de tongpositie bij articulatie van de klinker in de stam van het woord (van Putten, 2014). Daarnaast is in Tabel 1 te zien dat niet alleen vocalen, maar ook de consonant [l] een aantal keer in hoofdletters staat gedrukt. Deze klank kan namelijk een proces van nasalisatie ondergaan en worden uitgesproken als een [n], indien de stam van het woord een nasale consonant [m, n, ɲ, ŋ, ŋʷ] bevat (van Putten, 2014).

1.2 Semantiek in naamwoordsklassen

Nominale classificatiesystemen brengen categorisering van naamwoorden tot stand (Craig, 1986; Hendrikse, 2001). Maar wat die categorisering omvat, is taalafhankelijk en daarnaast niet altijd even duidelijk te achterhalen, mede door het feit dat nominale

classificatiesystemen diep in de grammatica van een taal geworteld zitten (Aikhenvald, 2000). Wel is er de laatste jaren meer consensus ontstaan over de rol die semantiek speelt in de classificatie van naamwoorden; deze systemen worden gezien als een grammaticale weergave van de categorieën die men cognitief maakt op basis van ervaringen (Sagna, 2012).

Het blijkt echter een lastige opgave om de semantische basis van naamwoordsklassen te achterhalen en deze probleemstelling wordt door onderzoekers op verschillende manieren benaderd. Er wordt bijvoorbeeld regelmatig gewerkt vanuit de *prototype theory* (Sagna, 2012), waarin de naamwoorden binnen semantische domeinen aan elkaar worden gelinkt middels polysemie, metonymie en metaforische extensie (e.g. Breedveld, 1995; Selvik, 2001). Gevonden patronen vanuit deze theorie lopen erg uiteen, omdat een onderzoeker er met deze werkwijze redelijk vrij in is zelf semantische domeinen te construeren. Bovendien kunnen semantische domeinen cultuurspecifiek zijn.

Een andere, bredere wijze van aanpak, is te kijken naar algemenere eigenschappen die referenten van woorden in eenzelfde klasse met elkaar gemeen kunnen hebben. Een voorbeeld van een dergelijke eigenschap, die in meerdere onderzoeken naar voren komt als een belangrijke factor in nominale classificatie, is *animacy*, ofwel levendheid. Dit semantische kenmerk verwijst naar een eigenschap van de referent van een woord; in welke mate leeft en voelt deze referent? Een woord verwijzend naar een mens staat bijvoorbeeld hoog in deze hiërarchie van levendheid, een woord voor een tafel staat laag. Van dit semantische kenmerk is volgens Croft (1994) wijd bekend dat het een rol speelt in naamwoordsclassificatie van talen, in ieder geval in het ontstaan ervan – een standpunt dat door meerdere onderzoekers wordt gedeeld (e.g. Di Garbo, 2013; Dixon, 1986).

Ook in het gros van onderzochte talen uit de Niger-Congo familie komt de rol van levendheid tot uitdrukking in de classificatie, hoewel de wijze waarop per taal kan verschillen. In een aantal talen komen mensen en dieren in dezelfde naamwoordsklasse voor (e.g. Carlson, 2011; Wit, 2015); in andere talen uit deze familie worden woorden voor menselijke referenten apart gecategoriseerd, terwijl dieren en objecten samen in klassen voorkomen (e.g. Childs, 2011; Contini-Morava, 1994; Morrison, 2011). De hierboven geciteerde publicaties zijn geschreven over Niger-Congo talen, de overkoepelende talenfamilie van het Avatime. Over de Kwa-familie die hieronder valt, kan echter weinig worden gevonden wat betreft de rol van levendheid in naamwoordsklassen. Harley (2005) geeft in zijn beschrijvende grammatica van de Kwa-taal Tuwuli (Niger-Congo, Ghana) wel een overzicht van

semantische overeenkomsten van de nomina in hun klassen, maar hierin is geen invloed van levendheid terug te zien.

1.3 Semantiek in Avatime naamwoordsklassen

Er is door Schuh (1995) eerder onderzoek uitgevoerd naar de semantiek van de Avatime naamwoordsklassen. Hij concludeerde dat vrijwel alle woorden voor mensen in klasse 1 te vinden zijn, dat veel stofnamen in klassen 5 en 7 zitten, dat tijdsaanduidingen veelal in klasse 3 te vinden zijn en dat er geen correlaties te vinden zijn tussen naamwoordsklasse en semantische categorieën als dieren, lichaamsdelen en objecten. Om te onderzoeken of deze bevindingen ook met een andere Avatime woordenset gevonden kunnen worden en om na te gaan of er eventueel nog andere patronen in de taal te vinden zijn, zoals die in andere talen naar voren kwamen, heb ik een kleinschalig corpusonderzoek uitgevoerd met een lijst van 592 Avatime woorden, beschikbaar gesteld door Saskia van Putten.

Uit de resultaten bleek de levendheid van referenten van groter belang te zijn dan fysieke kenmerken. Zo werd een vondst van Schuh (1995) bevestigd: naamwoorden die naar mensen refereren, komen vrijwel altijd in de eerste naamwoordsklasse voor. Daarnaast zijn er in deze klasse geen woorden voor niet-levende referenten gevonden. Verder blijken er naamwoordsklassen te zijn waarin juist helemaal geen woorden voor levende referenten voorkomen: klassen 4, 5 en 7. Hoewel dit onderzoek op een relatief kleine selectie woorden is uitgevoerd, lijken de resultaten erop te wijzen dat toewijzing van klassen aan naamwoorden niet geheel arbitrair is.

1.4 Huidig onderzoek

Er zijn voor verschillende talen resultaten gerapporteerd van onderzoeken naar semantische kenmerken in nominale classificatiesystemen. Het grootste deel van deze studies is gebaseerd op de kaders zoals die door de onderzoekers zijn opgesteld; er is weinig onderzoek gedaan in experimentele setting. Hierdoor wordt er aan het taalgevoel van de taalgemeenschap voorbij gegaan, terwijl de intuïtie van sprekers veel kan zeggen over de werking van de classificatie. Het biedt immers inzicht in de huidige staat van de grammatica; het laat zien of gevonden patronen nog een actief onderdeel zijn van het taalgebruik of slechts een overblijfsel van een systeem dat de taal vroeger kende.

Om meer informatie te verkrijgen over dit actievere aspect van nominale classificatie en de intuïties die Avatime sprekers hierover hebben, zal er in dit artikel worden voortgeborduurd op de voorgaande corpusstudie in de vorm van experimenteel onderzoek. Deze opzet is gebaseerd op de studie van Sagna (2012), die de semantiek achter de Gújjolaay Eegimaa naamwoordsklassen onderzocht door pseudowoorden aan sprekers aan te bieden in experimentele setting. In dit artikel zal de focus liggen op een factor waarvan enig effect te zien was in het corpusonderzoek: levendheid. Deze factor is, evenals de factor fysieke vorm, meegenomen in de opzet van Experiment 1. Op de rol van fysieke kenmerken zal in het huidige artikel echter niet worden ingegaan. Omdat er de meeste aanleiding was aan te nemen dat levendheid van belang is in het nominale classificatiesysteem van het Avatime, is ervoor gekozen om deze factor individueel nader te bestuderen in Experiment 2.

Voor Experiment 1 van deze studie zijn de deelnemers gevraagd uit zes afbeeldingen het plaatje te kiezen dat volgens hen het beste paste bij het pseudoword dat ze auditief kregen aangeboden. In Experiment 2 kregen participanten de opdracht uit twee gehoorde pseudowoorden het woord te kiezen dat volgens hen het best paste bij de afbeelding die ze te zien kregen. De gebruikte materialen en gehanteerde procedures van de twee experimenten zullen in de methodesectie in meer detail worden behandeld. De resultaten van Experimenten 1 en 2 zullen worden geanalyseerd en geïnterpreteerd om antwoord te kunnen geven op de volgende onderzoeksvraag: in hoeverre speelt de levendheid van referenten een rol in de actieve classificatie van naamwoorden door sprekers van het Avatime? Op basis van de gegevens van andere Niger-Congo talen en de resultaten van de corpusstudie, wordt er verwacht dat de levendheid van referenten invloed heeft op de associaties die men maakt tussen deze referenten en de aangeboden pseudowoorden. De hypothese is dat er meer plaatjes van levende referenten worden gekoppeld aan pseudowoorden uit de eerste klasse, en meer levenloze referenten aan pseudowoorden uit de vierde en vijfde klasse.

2. Experiment 1

2.1 Methode

2.1.1 Participanten

Experiment 1 is door twintig participanten uitgevoerd, van wie 9 vrouwen en 11 mannen. Hun leeftijden lagen tussen de 18 en 54 jaar ($M = 25.75$; $SD = 9.53$). Er is voornamelijk gekozen voor jonge personen, omdat het experiment wat ingewikkeld

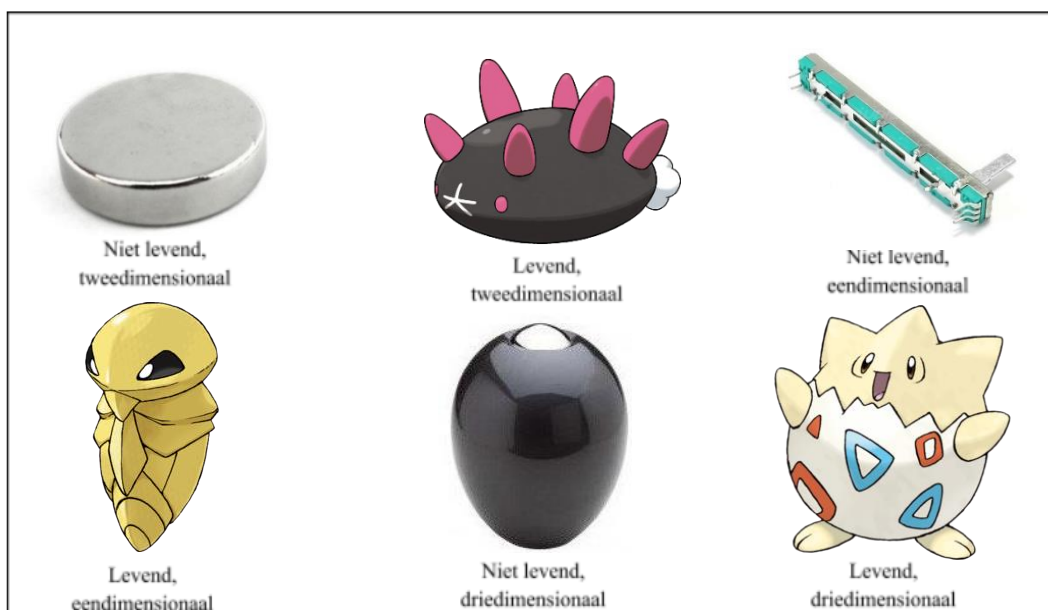
bleek voor oudere deelnemers. Deelnemers ontvingen een beloning van drie Ghanese cedi's. Aan de hand van de koers op dat moment, kan dit worden omgerekend naar ongeveer vijftig eurocent.

2.1.2 Materiaal

Afbeeldingen

In zowel Experiment 1 als Experiment 2 zijn er in totaal zestig verschillende afbeeldingen getoond aan de participanten, dertig van levende referenten en dertig van niet-levende. Binnen deze twee categorieën was er daarnaast nog het onderscheid tussen de drie verschillende vormen die zijn onderzocht in Experiment 1; eendimensionaal, tweedimensionaal en driedimensionaal, volgens de definities van Aikhenvald (2000) en Seifart (2005). De levenloze objecten waren bijvoorbeeld machineonderdelen, de levende referenten waren afkomstig uit het videospel Pokémon (Nintendo, z.d.). De zestig afbeeldingen vormden de uiteindelijke selectie die overbleef na een pre-test met drie sprekers van het Avatime, die werden gevraagd of ze een woord kenden voor de referenten op de plaatjes.

Voor Experiment 1 zijn deze afbeeldingen onderverdeeld in tien sets van zes afbeeldingen: drie levend, drie niet levend, twee van iedere vorm. Deze onderverdeling van plaatjes in sets is op willekeurige basis gemaakt. Een voorbeeld van een dergelijke afbeeldingenset is hieronder gegeven in Figuur 1.



Figuur 1. Een van de tien sets afbeeldingen zoals die, zonder onderschriften, aan de participanten zijn aangeboden in Experiment 1. De overige sets staan in de Appendix.

Pseudowoorden

Voor het onderzoek zijn er zestig pseudowoorden geconstrueerd om auditief aan de participanten aan te bieden. Deze woorden bestonden uit een prefix, een stam en een definitief lidwoord in de vorm van een suffix. Het suffix is evenals het prefix afhankelijk van de naamwoordsklasse en is toegevoegd aan de pseudowoorden om de klasse voor de participanten zo duidelijk mogelijk te maken. De prefixen van klasse 1 en klasse 2 zijn namelijk een minimaal paar, enkel verschillend in toonhoogte, terwijl de suffixen meer van elkaar verschillen.

De pseudowoorden zijn geconstrueerd aan de hand van een aantal fonologische regels. Twee van deze regels zijn reeds behandeld in de inleiding: de klinkerharmonie en de nasalisering van de [l]. De laatste regel van de Avatime fonologie die is gehanteerd bij het construeren van de pseudowoorden betreft de complementaire distributie van de klanken [l] en [r] zoals beschreven door Schuh (1995).

De pseudostammen die op basis van bovenstaande fonologie waren bedacht, zijn nog voorgelegd aan twee Avatime sprekers om na te gaan of ze niet voorkwamen in bestaande woorden. Deze pre-tests resulteerden in tien stammen, waar vervolgens zestig pseudowoorden mee zijn gevormd die de sprekers onbekend waren en die daardoor konden worden gebruikt in de studie. Een overzicht van de in Experiment 1 gebruikte pseudowoorden is te zien in Tabel 2. Het waren er zestig, om iedere set van zes afbeeldingen met een eigen pseudowoord te kunnen presenteren. De uiteindelijk geselecteerde pseudowoorden zijn ingesproken door Saskia van Putten en als audiofragmenten aan de participanten aangeboden.

Tabel 2

Overzicht van de zestig pseudowoorden zoals gebruikt in Experiment 1.

	<i>Klasse 1</i>	<i>Klasse 2</i>	<i>Klasse 3</i>	<i>Klasse 4</i>	<i>Klasse 5</i>	<i>Klasse 6</i>
<i>Be</i>	obeye	òbelo	libele	kibeye	kubeo	kebea
<i>Blu</i>	obluye	òblulo	liblule	kibluye	kublulo	keblua
<i>Dre</i>	òdreye	òdrelò	lìdrelè	kìdreye	kùdrelò	kadrea
<i>Fe</i>	ofeye	òfelo	lifele	kifeye	kufeo	kefea
<i>Ge</i>	ogeye	ògelo	ligele	kigeye	kugeo	kegea
<i>Klo</i>	okloyè	òklolo	liklole	kikloye	kuklo	kekloa
<i>Kpe</i>	òkpèye	òkpelò	lìkpelè	kìkpeye	kùkpelò	kakpèa
<i>Mwò</i>	òmwòye	òmwòno	lìmwòne	kìmwòye	kùmwò	kamwòa
<i>Nu</i>	onuye	ònuno	linune	kinuye	kunuo	kenua
<i>Pla</i>	òplaye	òplalò	lìplalè	kìplaye	kùplalò	kapla

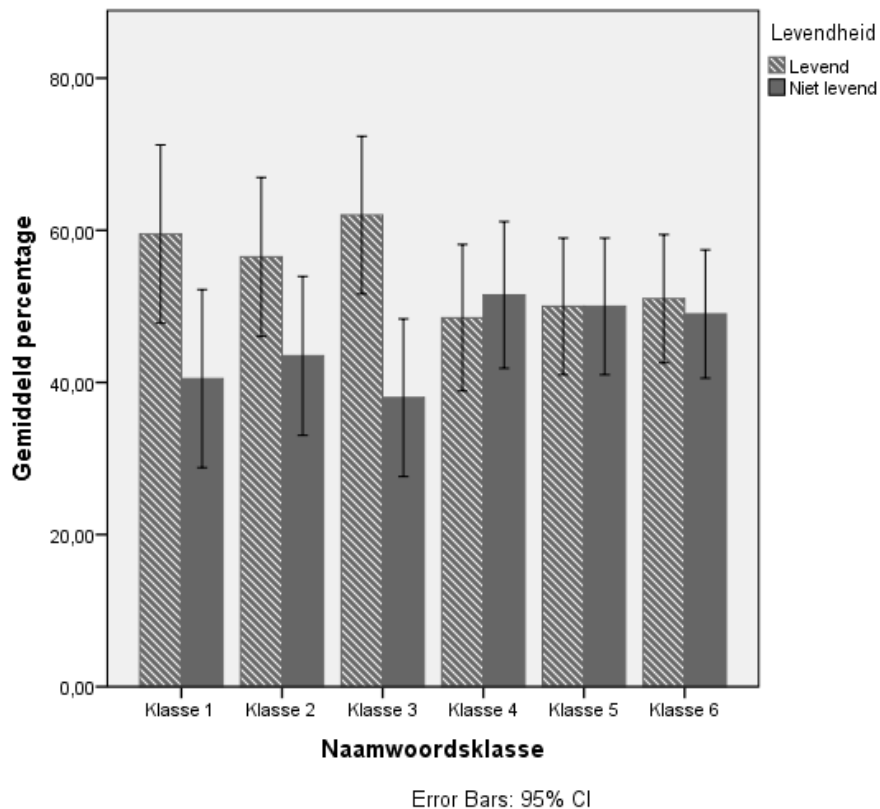
2.1.3 Procedure

In Experiment 1 kregen de participanten via een PowerPointpresentatie de zestig audiofragmenten met pseudowoorden een voor een aangeboden. Het merendeel van de participanten heeft deze woorden via een koptelefoon gehoord, voor de eerste acht deelnemers was deze nog niet beschikbaar. Terwijl de participanten een pseudowoord hoorden, werd er op het beeldscherm van een laptop een set van zes afbeeldingen getoond. De stam van ieder pseudowoord is in het experiment zesmaal aan bod gekomen; telkens met dezelfde set van zes afbeeldingen, maar met een ander pre- en suffix. Op deze wijze zijn in het experiment alle combinaties van prefixen en sets van afbeeldingen aan bod gekomen. Welke pseudowoorden er gelijktijdig met welke afbeeldingensets zijn aangeboden, is willekeurig bepaald. Een overzicht van deze sets en bijbehorende pseudowoorden is gegeven in de Appendix.

De participanten werden gevraagd zich voor te stellen dat het gehoorde pseudowoord daadwerkelijk zou bestaan en de afbeelding aan te wijzen die er, naar hun idee, het beste mee beschreven zou kunnen worden. Hieraan werd toegevoegd dat ze af moesten gaan op hetgeen dat het eerst in hen op zou komen. Deze instructies werden gegeven in het Engels en zin voor zin door een tweetalig persoon nogmaals herhaald in het Avatime, ook voor participanten met een goede Engelse taalvaardigheid. De antwoorden van participanten werden direct op een tweede laptop bijgehouden.

2.2 Resultaten en discussie

Voor de resultatenanalyses van dit experiment zijn er voor iedere deelnemer scores in percentages gehanteerd. Er is gewerkt met scores van de aantallen levende referenten die participanten kozen, met zes scores voor iedere participant: een aparte score voor iedere naamwoordsklasse waaruit pseudowoorden werden aangeboden. De gemiddelde percentages gekozen afbeeldingen van levende referenten en bijbehorende standaarddeviaties staan schematisch weergegeven in Figuur 2.



Figuur 2. Percentages gekozen levende en levenloze referenten bij het auditieve aanbod van pseudowoorden in verschillende naamwoordsklassen.

Om het effect te onderzoeken van de naamwoordsklasse van een auditief aangeboden pseudoword op het aantal gekozen levende referenten als best passende afbeelding bij pseudowoorden met een prefix van deze naamwoordsklasse, is er een *repeated measures* ANOVA uitgevoerd. Uit de toets bleek dat de keuze voor afbeeldingen van levende referenten niet significant werd beïnvloed door de naamwoordsklasse waar het gehoorde woord toe behoort, $F(5, 95) = 2.17$, $p = .064$, $\eta^2_p = .102$. Geplande contrasten toonden over het algemeen geen significante verschillen in de percentages levende referenten tussen de zes naamwoordsklassen, behalve tussen klassen 3 en 4, $F(1, 19) = 6.75$; $p = .018$; $\eta^2_p = .262$.

Uit de resultaten van Experiment 1 blijkt dat participanten afbeeldingen van levende referenten over het algemeen verkozen boven die van levenloze referenten. In Figuur 2 is te zien dat deze voorkeur met name van toepassing was wanneer participanten afbeeldingen kozen voor pseudowoorden uit de eerste, tweede en derde naamwoordsklasse. Deze resultaten komen, wat betreft de eerste naamwoordsklasse, overeen met de bevindingen van Schuh (1995) en het corpusonderzoek. Verder bleek klasse 4 de enige naamwoordsklasse waarbij participanten de voorkeur hadden voor afbeeldingen van levenloze referenten en bij klasse 5 zijn de percentages gelijk. Met deze scores lijken deze twee naamwoordsklassen te verschillen van de eerste drie, al

is het verschil met de percentages bij de zesde naamwoordsklasse, tegen de verwachtingen in, niet erg groot. Statistische toetsing toonde aan dat deze bevindingen niet genoeg aanleiding zijn om de hypothese, dat er meer levende referenten worden gekozen bij een woord uit klasse 1 en meer levenloze bij een woord uit klasse 4, aan te nemen.

3. Experiment 2

3.1 Methode

3.1.1 Participanten

Aan Experiment 2 hebben ook twintig personen deelgenomen, die geen participant van het eerste experiment waren geweest. Ze waren tussen de 18 en 70 jaar ($M=40.45$; $SD=19.33$). De verhouding mannen en vrouwen was in deze groep gelijk. Voor deelname aan dit experiment gold eveneens een beloning van drie Ghanese cedi's.

3.1.2 Materiaal

Voor Experiment 2 zijn dezelfde afbeeldingen gebruikt als voor Experiment 1. Een verschil tussen de experimenten, is dat er bij Experiment 2 voor is gekozen om slechts vijf pseudostammen in het experiment op te nemen (*blu*, *drɛ*, *klo*, *mwɔ* en *pla*) om het design en de analyse simpel te kunnen houden. De keuze voor deze vijf gebruikte stammen was willekeurig.

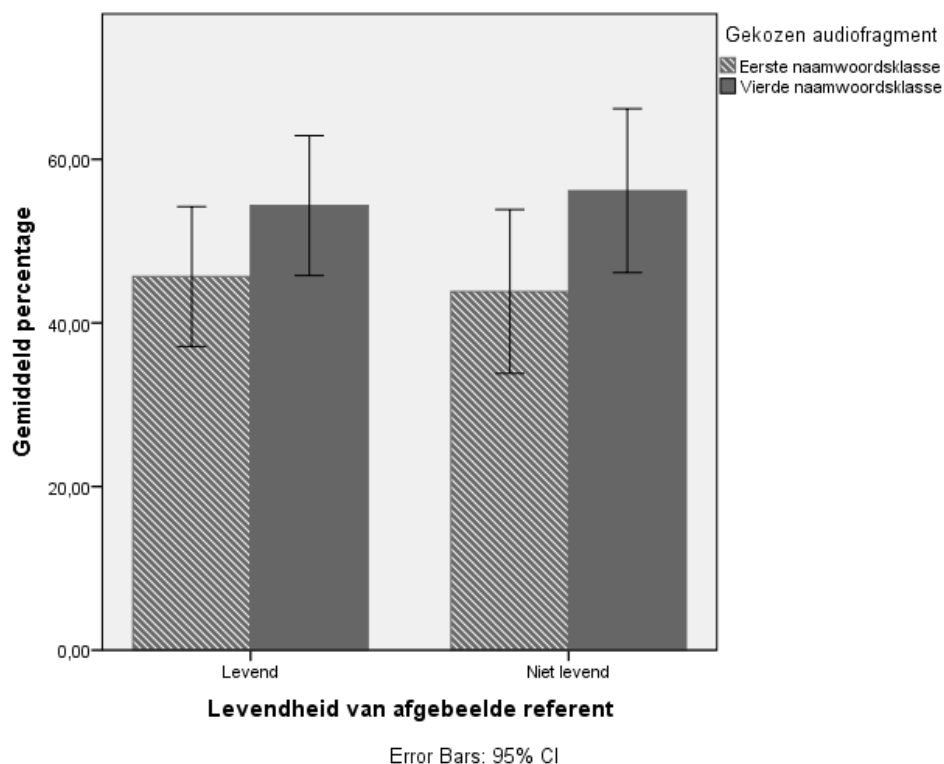
3.1.3 Procedure

In Experiment 2 zijn alle zestig afbeeldingen een voor een aan de participanten getoond met behulp van een PowerPointpresentatie. Bij de presentatie van een afbeelding hoorden ze twee keer dezelfde pseudostam, maar de woorden verschilden van elkaar in prefix en suffix. Er werd de participanten gevraagd het best passende woord te kiezen en te gaan voor het audiofragment dat hun in eerste instantie het beste leek. Van de twee pseudowoorden die de deelnemers hoorden in een item, kwam er één woord uit klasse 1 en één woord uit klasse 4. Er is voor deze klassen gekozen met de reden dat klasse 1 de enige klasse is die alleen maar woorden bevat die naar levende referenten verwijzen, en klasse 4 daar tegenover staat als een van de twee klassen

waarin juist geen enkel woord voor een levende referent zit. In het experiment zijn er vijf pseudostammen (en dus tien pseudowoorden) aan de participanten aangeboden. Welke pseudowoorden er gelijktijdig met welke afbeeldingen zijn aangeboden, is willekeurig bepaald. Een overzicht hiervan is gegeven in de Appendix. Ook bij dit experiment zijn alle participanten in zowel het Engels als het Avatime geïnstrueerd en werden de antwoorden genoteerd op een tweede laptop.

3.2 Resultaten en discussie

In Experiment 2 werd onderzocht of de levendheid van de afgebeelde referent effect heeft op het pseudowoord dat men kiest als beste match bij de afbeelding. De statistische analyse van dit experiment is, net als die van Experiment 1, uitgevoerd met percentages als afhankelijke variabele. Er is voor iedere participant bijgehouden hoeveel procent van de items, bestaande uit een afbeelding van ofwel een levende ofwel een niet levende referent, zij hebben gekoppeld aan een pseudowoord uit de eerste naamwoordsklasse en hoeveel procent van de items aan een pseudowoord uit de vierde naamwoordsklasse. Deze resultaten zijn afgebeeld in Figuur 3.



Figuur 3. Percentages gekozen pseudowoorden uit de eerste en vierde naamwoordsklasse bij aanbod van afbeeldingen van levende en niet-levende referenten.

De resultaten van Experiment 2 zijn geanalyseerd aan de hand van een *repeated measures* ANOVA. De ANOVA wees uit dat de levendheid van de referent op de getoonde afbeelding geen significant effect had op de percentages gekozen audiofragmenten van pseudowoorden uit de eerste naamwoordsklasse, $F(1, 19) = .192$; $p = .666$; $\eta^2_p = .01$.

Wat betreft de resultaten van Experiment 2, is in Figuur 3 te zien dat participanten gemiddeld vaker kozen voor de audiofragmenten van pseudowoorden uit de vierde naamwoordsklasse, ongeacht de levendheid van de referent die op het plaatje werd afgebeeld. Dit komt niet overeen met de verwachtingen voor dit onderzoek. Ook de statistiek die voor dit experiment is uitgevoerd toonde geen significante effecten, waardoor er niet kan worden aangenomen dat een levende referent vaker wordt gekoppeld aan een woord uit klasse 1, en een levenloze vaker aan een woord uit klasse 4.

4. Algemene discussie

In dit onderzoek is er getracht antwoord te vinden op de vraag: in hoeverre speelt de levendheid van referenten een rol in de actieve classificatie van naamwoorden door sprekers van het Avatime? In meerdere onderzoeken naar talen in de Niger-Congo talenfamilie kwam levendheid regelmatig terug als een belangrijke factor in de nominale classificatiesystemen (e.g. Carlson, 2011; Childs, 2011). Om na te gaan of er vergelijkbare patronen in het Avatime voorkomen, is er voorafgaand aan dit onderzoek een kleinschalig corpusonderzoek gedaan naar de rollen van semantische (en fysieke) kenmerken in de naamwoordsklassen van deze taal. In het huidige onderzoek ben ik verder gegaan met deze resultaten om meer duidelijkheid kunnen te creëren over de naamwoordsclassificatie van het Avatime. Voorafgaand aan het onderzoek werd er verwacht een effect te vinden van levendheid op de nominale classificatie in de taal. De hypothese is getoetst aan de hand van twee experimenten, die beide zijn afgenomen bij een andere groep van twintig moedertaalsprekers van het Avatime.

De resultaten van beide experimenten in dit onderzoek bleken niet significant te zijn, maar waar de uitkomsten van Experiment 2 überhaupt niet overeenkwamen met de hypothesen, leek er in de resultaten van Experiment 1 wel iets van de verwachtingen terug te komen. Participanten leken hier levende referenten boven niet-levende te verkiezen wanneer ze woorden hoorden uit de eerste, tweede en derde naamwoordsklasse, maar niet bij woorden uit klassen 4, 5 en 6. Het is onduidelijk waar

het verschil in resultaten van de twee experimenten door verklaard kan worden. Het is mogelijk dat het design van een van de experimenten niet geschikt was om het huidige onderzoek mee uit te voeren, maar dit kan op basis van deze uitkomsten niet met zekerheid worden gesteld.

Het feit dat geen van de analyses wijzen op een significante relatie tussen levendheid van referenten en de naamwoordsclassificatie kan door een aantal factoren worden verklaard. Allereerst is het denkbaar dat de antwoorden van de eerste acht participanten minder betrouwbaar zijn door het feit dat zij de audiofragmenten zonder koptelefoon kregen aangeboden en de audiofragmenten daardoor niet goed genoeg konden verstaan. Een mogelijk waarschijnlijker verklaring voor de resultaten, echter, zit in de gebruikte items. In de testfase van de stimuli is er namelijk niet aan Avatime sprekers gevraagd of de referenten op de afbeeldingen levend of levenloos waren. Hoewel het erop lijkt dat veel participanten alsnog enig onderscheid maakten in levendheid, is het mogelijk dat het niet voor alle deelnemers even duidelijk was dat de plaatjes van Pokémon levende referenten voor moesten stellen. Zeker in de context van Experiment 2, waarin alle afbeeldingen geïsoleerd werden aangeboden en dus niet met elkaar konden worden vergeleken, is het in te beelden dat het lastig kan zijn op deze manier de levendheid van een referent te bepalen.

Wellicht had een andere opzet, waarin participanten de tijd hadden gehad om te reflecteren over hun taal en hun eigen taalgebruik, geresulteerd in andere conclusies. De hoge mate waarin een systeem van naamwoordsclassificatie is gegrammaticaliseerd (Aikhenvald, 2000) is een mogelijke reden dat het voor participanten lastig is om hun kennis hierover direct toe te passen, zeker in een abstracte setting als de experimenten waaraan zij deelnamen.

5. Conclusie

In Figuur 2, waarin de resultaten van Experiment 1 worden weergegeven, is te zien dat de participanten op een andere manier omgingen met levende en niet-levende referenten als het gaat om hun toewijzing aan de zes onderzochte naamwoordsklassen. We kunnen hier echter niet spreken van duidelijke verschillen tussen de klassen, omdat de statistische analyses geen significante effecten aantoonde. Ook een analyse van de data van Experiment 2 resulteerde niet in significante uitkomsten. Hiermee kan de onderzoeksvraag (In hoeverre speelt de levendheid van referenten een rol in de actieve classificatie van naamwoorden door sprekers van het Avatime?) als volgt worden beantwoord: hoewel de participanten in een van de experimenten enig taalgevoel leken te hebben over de semantiek achter

nominale classificatie, kan er, vanwege de niet-significante uitkomsten van het onderzoek, nog niet met zekerheid worden gesproken van een belangrijke rol van deze semantiek.

Er zijn nog genoeg mogelijkheden voor vervolgonderzoek naar de rol van semantische en fysieke kenmerken in de naamwoordsclassificatie van het Avatime, en dan met name naar de intuïtie die sprekers hierover hebben. Er zouden nog experimentele studies kunnen worden uitgevoerd met eenzelfde opzet als die van dit onderzoek, waarbij behandelde verbeterpunten in acht dienen te worden genomen. Daarnaast zou er nog een corpusstudie uit kunnen worden gevoerd waarbij niet de onderzoeker, maar sprekers van het Avatime het semantisch domein en de vorm van voorwerpen bepalen, voor zover dit mogelijk is. Het huidige onderzoek kan als basis worden gebruikt in vervolgstudies, omdat er een completer beeld mee is geschetst van de Avatime taal en de manier waarop deze door haar sprekers wordt gebruikt.

Referenties

- Aikhenvald, A. (2000). *Classifiers: A typology of noun categorization devices*. Oxford: Oxford University Press.
- Blench, R. (2006). A comparative study of the Ghana-Togo Mountain languages. Geraadpleegd op 29 juli 2019.
- Breedveld, J.O. (1995). The semantic basis of noun class systems. The case of the ki and nge classes in Fulfulde. *Journal of West African Languages*, 25(2), 63-74.
- Carlson, R. (1994). *A grammar of Supyire*. Berlijn, Duitsland: Mouton de Gruyter.
- Childs, G.T. (2011). *A grammar of Mani* (Vol. 54). Berlijn, Duitsland: Walter de Gruyter.
- Contini-Morava, E. (1994). *Noun classification in Swahili*. (Dissertatie). New York, NY: Columbia University. Geraadpleegd van <http://www2.iath.virginia.edu/swahili/oldversion/swahili.html>.
- Craig, C.G. (Red.). (1986). *Noun classes and categorization: Proceedings of a symposium on categorization and noun classification*. Amsterdam, Nederland: John Benjamins.
- Croft, W. (1994). Semantic universals in classifier systems. *Word*, 45(2), 145-171.
- Di Garbo, F. (2013). Evaluative morphology and noun classification: a cross-linguistic study of Africa. *SKASE Journal of Theoretical Linguistics*, 10(1), 114-136.
- Dixon, R.M. (1986). Noun classes and noun classification in typological perspective. In C.G. Craig (Red.), *Noun Classes and Categorization*, 105-112. Amsterdam, Nederland: John Benjamins.
- Fell, J.R. (1918). *A Tonga grammar*. Londen, Engeland: Society for promoting Christian knowledge.
- Hammarström, H., Forkel, R., & Haspelmath, M. (2019). *Glottolog 4.0*. Max Planck Institute for the Science of Human History. Jena, Duitsland. Geraadpleegd op 29 juli 2019, van <https://glottolog.org/>.
- Harley, M.W. (2005). *A descriptive grammar of Tuwuli, a Kwa language of Ghana*. (Dissertatie). University of London.
- Hendrikse, A.P. (2001). Systemic polysemy in the Southern Bantu noun class system. In H. Cuyckens & B.E. Zawada (Red.), *Polysemy in Cognitive Linguistics*, 185-212. Amsterdam, Nederland: John Benjamins.
- Lewis, P.M., Simons, G.F., & Fennig, C.D. (Red.). (2013). *Ethnologue: Languages of the World, Seventeenth Edition. Online version*. Dallas, Texas, TX: SIL International. Geraadpleegd van <http://www.ethnologue.com/>.
- Morrison, M.E. (2011). *A reference grammar of Bena*. (Dissertatie). Houston, TX: Rice University.
- Nintendo. (z.d.). [Informatie over Pokémon]. Geraadpleegd op 28 mei 2019 van <https://www.pokemon.com/us/pokedex/>.

- Sagna, S. (2012). Physical properties and culture-specific factors as principles of semantic categorization of the Gújjolaay Eegima noun class systems. *Cognitive Linguistics*, 23, 129-163.
- Seifart, F. (2005). *The structure and use of shape-based noun classes in Miraña (North West Amazon)*. (Dissertatie). Nijmegen: Radboud Universiteit Nijmegen.
- Schuh, R.G. (1995). Aspects of Avatime phonology. *Studies in African Linguistics*, 24(1), 31-67.
- Selvik, K.A. (2001). When a dance resembles a tree. In H. Cuyckens & B.E. Zawada (Red.), *Polysemy in Cognitive Linguistics*, 161-184. Amsterdam, Nederland: John Benjamins.
- van Putten, S. (2014). *Information structure in Avatime*. (Dissertatie). Nijmegen: Radboud Universiteit Nijmegen.
- van Putten, S. (2015). *Les 5: Nominale constituenten* [Ongepubliceerde handout].
- Wit, G.D. (2015). *Liko phonology and grammar: a Bantu language of the Democratic Republic of the Congo*. (Dissertatie). Utrecht: LOT.

Appendix

Materialen

Experiment 1

In dit deel van de Appendix staan de tien verschillende sets van zes afbeeldingen gegeven met voor iedere set de pseudostam waarmee deze telkens aan de participanten werd gepresenteerd.



Pseudostam: *be*

Pseudostam: *nu*



Pseudostam: *blu*

Pseudostam: *klo*



Pseudostam: *mwo*

Pseudostam: *pla*



Pseudostam: *dre*

Pseudostam: *fe*

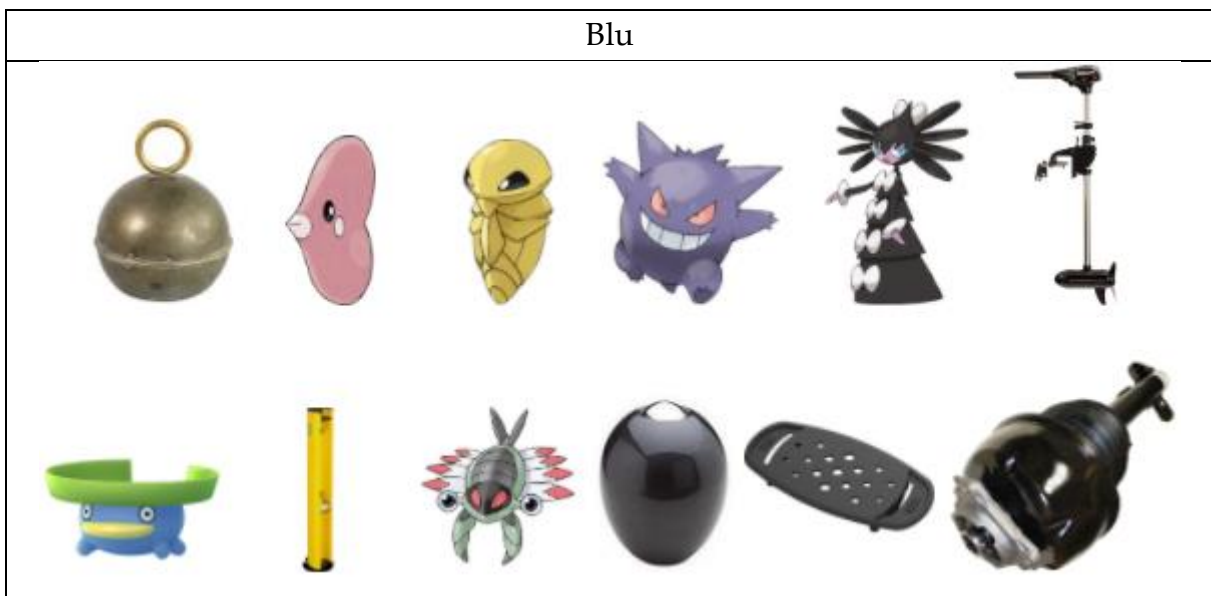


Pseudostam: *kpe*

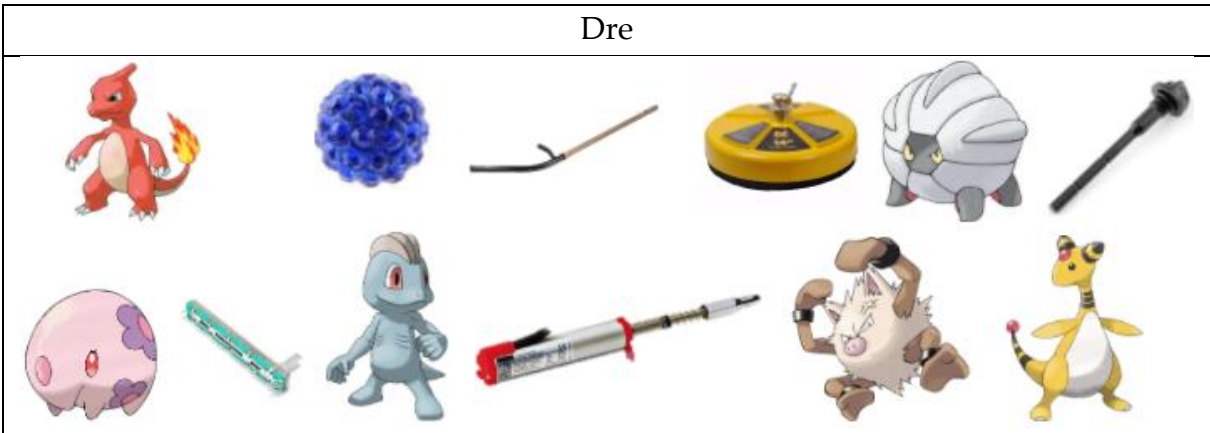
Pseudostam: *ge*

Experiment 2

In dit deel van de Appendix staan de aangeboden pseudowoorden gegeven met voor ieder woord de twaalf afbeeldingen die er gelijktijdig mee zijn gepresenteerd.



Dre



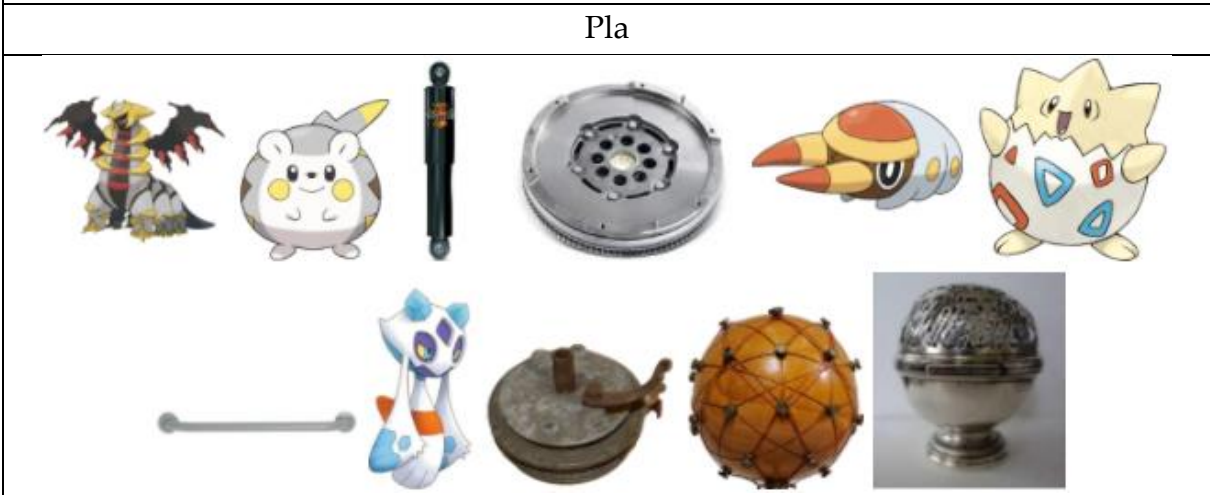
Klo



Mwo



Pla



Dankwoord

Het uitvoeren van dit onderzoek was niet mogelijk geweest zonder mijn begeleider, dr. Saskia van Putten, die mij gedurende het hele proces altijd hulp kon bieden wanneer nodig. Ik wil haar daarnaast, evenals tweede lezer prof. dr. Helen de Hoop, graag bedanken voor de beoordeling van mijn scriptie. Ook wil ik graag Radboud Universiteit Nijmegen bedanken voor het beschikbaar stellen van de financiële middelen om het onderzoek in Ghana plaats te kunnen laten vinden. Verder wil ik mijn peer reviewers bedanken voor het helpen verbeteren van dit artikel. Tot slot zijn er veel mensen in Vane die mij hebben geholpen door deel te nemen aan de experimenten of het onderzoek op een andere manier te ondersteunen, dus ook voor hen: *mlɔ mlɛwɔ lixwɛ!*

Language relativity re-visited:

Perception of blue and green in Greek, Irish, and German¹

M.C.A. (Mignon) van Vorselen

Master *General Linguistics*, Radboud University Nijmegen

Abstract: Some languages have multiple terms for one colour category. For example, Greek has two terms for blue and Irish has two terms for green. Studies into this phenomenon are closely related to the Sapir-Whorf Hypothesis. This hypothesis states that language influences one's perception of the world. In academic literature, this hypothesis is mainly tested in relation to colour perception. Studies often establish that speakers of a language with multiple terms for one colour have a different perception of this colour and, therefore, a different perception of the world.

In this research proposal, adjustments to the research design of a study by Thierry et al. (2009) are proposed. Thierry et al. observed ERP-data retrieved from Greek and English participants participating in a recognition task. They concluded that perception differences between the two groups existed. The proposed design in this paper aims to observe ERP-data retrieved from Greek, Irish and German participants. The goal is to see if there are differences between languages having multiple terms for the colour blue or green, like Greek or Irish, and languages not making this distinction, like German. Furthermore, it is proposed to study if the expected differences in Greek- and Irish-speaking participants are similar to each other and if they are contrasting to the findings found in German-speaking participants.

The potential outcome of this proposed study is a difference in brain activity between languages with one term for one colour category and languages with multiple terms for one colour category. It is also expected that the brain activity in languages distinguishing two terms for blue (e.g. Greek) and the brain activity in languages distinguishing two terms for green (e.g. Irish) will be sharing the same characteristics and thus will be similar to each other.

Key words: language relativity, colour perception, Sapir-Whorf hypothesis, research proposal

¹ An adjustment of Thierry et al. (2009)

1. Introduction

For ages, researchers have been interested in the topic of how language influences one's thoughts. This can be seen in the ideas from Wilhelm von Humboldt, in fictional literature by the author George Orwell, and in the linguistic works of Edward Sapir and Benjamin Lee Whorf (Koerner, 1992). The Sapir-Whorf hypothesis is a principle that states that human thought is affected by language. According to the hypothesis, this results in speakers of different languages having a different view of the world (Regier & Yang, 2017). This hypothesis has two approaches, namely a strong approach and a weak approach. The strong approach states that language completely determines cognition. This means that one's thoughts are never free from the constraints that exist in the language that is spoken. The weak approach makes a less strong distinction. In this case, it is believed that language influences cognition in habitual thought patterns, categorising issues and retrieval of objects and events (Han & Cadierno, 2010).

The large number of research carried out over the years shows that perception of colour is one of the most ideal test cases for this hypothesis (Heurley, et al., 2012). Within most Germanic Indo-European languages, colours usually have one term (Kay & Maffi, 2013). In languages like English, German and Dutch, a further discrimination within a colour term is not obligatory. This means there is a need for just one word for blue, one word for green and so on. In other language families, this pattern is not always followed (Brown & Lenneberg, 1954). Greek, for instance, has two terms for what is called "blue" in English to distinguish between dark and light blue. Furthermore, Irish has two terms for what is called "green" in English, distinguishing between natural and artificial green. In both languages this distinction is mandatory (Coventry et al., 2006; Swinkels, 2015). According to the Sapir-Whorf hypothesis, it could be assumed that speakers of Greek and Irish perceive those colours (and thus the world) differently than speakers of a Germanic Indo-European languages like English, German or Dutch.

This assumption has been tested in the case of multiple languages with two terms for the English "blue", including Greek and Russian (Thierry et al., 2009; Winawer et al., 2007). Thierry et al. (2009) researched if Greek-speaking participants and English-speaking participants had a different electroencephalography (EEG) reaction to varying shades of blue. Participants were observed via EEG while performing a distinguishing task. The results of the study indeed show that there was a greater distinction between different shades of blue than different shades of green in Greek participants. In contrast, English speakers did not show this distinction. In another study, Winawer et al. (2007) researched if speaking Russian would lead to differences

in colour discrimination in comparison to English. In this study, Russian and English participants took part in a perceptual task in which they had to categorise shades of blue in different conditions. The results of this study also suggest that there is a difference between speakers of a language with multiple terms for one colour and speakers of English.

Although both studies seem to imply that having two terms for a colour indeed results in respectively different electrophysiological data and a category advantage, there are many gaps in the study designs resulting in a lot of uncertainty. For example, the Greek participant group lived in England, and the Russian participant group lived in the United States of America (Thierry et al., 2009; Winawer et al., 2007). Therefore, the influence of English could have been a factor that might have altered the results. An important aspect of conducting research is to ensure that the influence of external factors is kept to a minimum. In the case of colour perception within languages, it is important to look at which languages are spoken by the speakers and to what level this external factor affects the tested language. In colour perception research, it is studied whether there is a difference in colour perception between native speakers of two different languages. It is therefore important to establish with certainty that a participant in such an investigation also uses the language in question to distinguish the colours and that there is no influence from a possible second language.

In both studies mentioned above, only English was used as a control language. It can therefore not be ruled out that the effects shown in Thierry et al. (2009) and Winawer et al. (2007) are only present in the tested languages in comparison with English. Despite this, it could be concluded that, based on the earlier mentioned literature, one's (native) language, or more precisely, the number of terms for a colour, influences the perception of colour. However, what is not fully known is what this influence is, where this influence stems from and if it occurs only in languages distinguishing different blues. The design presented in Thierry et al. was a good design as it enables one to observe how the brain reacts to the appearance of a particular use with the use of EEG. If this is done for different speakers of multiple languages, it is possible to compare the event-related potential (ERP) results and see if there is a difference.

To rule out the noise, as mentioned earlier, it is important to look closely at the weaknesses in the study by Thierry et al. (2009). One of the biggest pitfalls in this study might be the choice of languages. One of the participant groups (the Greek group) is residing in England, therefore the influence of English cannot be excluded. Furthermore, there is no form of comparison with another language having a different colour distinction. As a result, coincidence cannot be completely ruled out. By looking

closely at the languages used in the research design, the noise can be cancelled out and the real influence of having multiple terms for a single colour can be examined. This would add the actual perception and impact of specific terms for colour to our understanding and show if there is a pattern within retrieving colours from the brain.

For this reason, this research proposal suggests a number of adjustments to the design of Thierry et al. (2009). Thierry et al. focused on the following research question: “Does the existence of two colour terms in Greek lead to greater and faster perceptual discrimination in native speakers of Greek than in native speakers of English?” In this study design, 20 native English speakers and 20 native Greek speakers were examined while they had to do a basic shape discrimination task. This task consisted of 4 blocks with 16 green or blue circles and a square. During this experiment, electrophysiological data was recorded. Thierry et al. concluded that the comparison of the electrophysiological data indeed shows that the Greek participants made a different distinction between the shades of blue than between the shades of green. English participants do not show any differences in distinction. Even though the research by Thierry et al. has been quoted frequently in other studies (Gallagher, 2017; Regier & Kay, 2009; Simmons et al, 2009) some adjustments for improvement of the design are brought forward in the design of this proposal.

The proposed design will focus on two questions. First, it will be examined if there is a significant difference of colour perception between a language having multiple terms for “blue”, a language having multiple terms for “green”, and a language with one term for each colour. Based on previous literature (Winawer et al. 2007; Thierry et al. 2009), this difference can be expected, at least for blues. If this difference is established, it is examined whether this recorded difference is comparable in EEG data for participants distinguishing blue and participants distinguishing green. It can be expected that both participant groups show certain overlapping differences in common properties, such as equivalent peak size of the brain potential amplitude, simultaneous intervals and firing of the same neurons at the same moment in time. However, because mainly languages with a distinction in terms of blue have been studied, it is not known whether the properties for blue can also be assumed for other colours, such as green. If this could be assumed it could be discussed if some sort of general practice for processing multiple colour terms exists.

Studying the differences in colour perception between languages is important because it can say something about the worldview of different language populations. With this study, the neural patterns of colour perception can be determined and it answers if the effects, as seen in the literature, can be generalised to multiple languages with the same phenomenon.

2. Method

2.1 Participants

The first adjustment concerns the participants of the study. Thierry et al. (2009) focussed on English and Greek participants. In this experiment, brain event-related potentials (ERPs) are recorded for monolingual speakers of Greek living in Greece, monolingual speakers of Irish living on the Irish islands, and monolingual speakers of German living in Germany or Austria. The addition of an Irish participant group to the set-up of Thierry et al. has been done so that a comparison can be made between the ERP data of distinguishing blue terms and distinguishing green terms (in other words, answering the second sub-question). Because an Irish participant group will be chosen, it is important to exclude the influence of other languages. On the Irish island (The Republic of Ireland and Northern Ireland) English is widely spoken in addition to Irish. In order to keep the differences between the groups of participants clear-cut, it is decided to exchange English for a language with the same number of terms for a colour. For this reason, it is decided to change the English participant group for a German participant group, as both English and German have the same colour classification within their language. Similar to Thierry et al., the same size of participant groups applies ($n=20$).

Due to the nature of the study, it is important to recruit monolingual participants. It is chosen to recruit participants within secondary vocational educational schools where English does not play a major role. Finding Greek monolingual participants and German monolingual participants is not expected to be a problem. However, given the influence of English within Ireland and Northern Ireland, it cannot be expected to find monolingual speakers of Irish. A solution might be to establish that Irish is the preferred language over English within the participant group. This can be done with a survey.

2.2 Material

Recruitment material is needed to recruit participants. In order to determine whether the participants are suitable for this study, it has to be determined whether they are monolingual Greek or German or whether they have Irish as their preferred language. A survey is used to retrieve language information on all the participants and an English proficiency test is used for the Greek and Irish participants to check if their proficiency in English is not too high, enabling it from interfering with their colour

perception in Greek or Irish. This is not necessary for the German participants as both English and German have the same amount of terms for one colour and interference is therefore not an issue.

During this experiment the stimuli presented in Figure 1 and Figure 2 are used to assess the participants' perception of the presented colours. As described earlier, Greek has two terms for blue, namely *ble* for darker blue and *ghalazio* for a lighter blue. Irish has two terms for green, namely *glas* for natural (darker) green and *uaine* for artificial (lighter) green. German only has one term for both categories, namely *blau* for all shades of blue and *grün* for all shades of green. The colour classification of the colours blue and green can be seen in Figure 1. The colours used in this experiment will be #6FDEFF for light blue, #0D3078 for dark blue, #94FB86 for light (artificial) green and #2D5F26 for dark (natural) green. For the measurement of brain responses, an ERP set-up is used.

	
Greek: <i>Ble</i> Irish: <i>Gorm</i> German: <i>Blau</i>	Greek: <i>Ghalazio</i> Irish: <i>Gorm</i> German: <i>Blau</i>
	
Greek: <i>Prásinos</i> Irish: <i>Glas</i> German: <i>Grün</i>	Greek: <i>Prásinos</i> Irish: <i>Uaine</i> German: <i>Grün</i>

Figure 1. Colour terms in Greek, Irish, and German.

2.3 Task

In order to establish that the participant groups do indeed determine the colours as shown in Figure 1, a test run with a test group must be carried out first. In this test run some speakers of Greek, Irish and German have to name the colours of the circles in Figure 1. These speakers will not participate in the experiment. Once the correctness of the colour terms has been established, the real experiment can take place.

All the participants are going to participate in a basic shape discrimination task (Figure 2) based on the stimuli sample of Thierry et al. (2009). In this task, participants have to look at the shapes presented in four different blocks while wearing an EEG cap to collect neurological data. The blocks will consist of a random pattern with 16 light or dark circles in respectively green or blue and one square in respectively green or blue. The shapes will be presented one by one for 200 milliseconds and the inter-

stimulus duration will be 800 milliseconds. The participants are asked to push a button when they observe a square within the stimuli pattern. This task is a distraction task to maintain the attention of the participants. While the participants complete the task, the EEG cap records brain responses to the stimuli below.

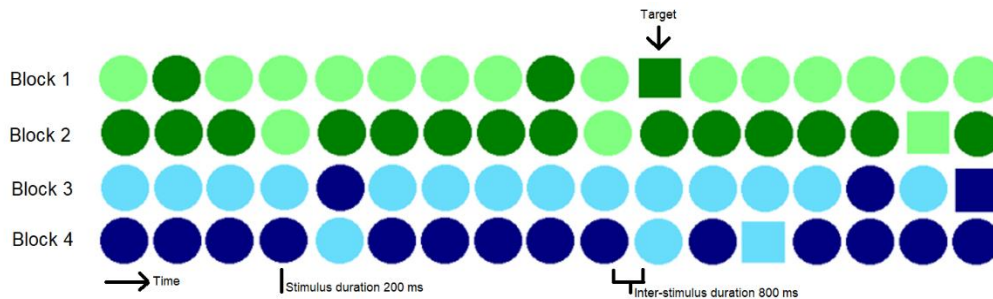


Figure 2. Stimuli sequences.

2.4 Analysis

After obtaining ERP data, this data will be filtered using, among other things, a band pass to create a clear graph from the brain activity. When the data is complete and edited, the data can be analysed. The first analysis looks at the differences in ERP data between the Greek and German group of participants and the differences in ERP data between the Irish and German groups of participants. This analysis will be used to try to answer the first question, namely “Is there a significant difference of colour perception between a language having multiple terms for blue, a language having multiple terms for green and a language with one term for each colour?”

The second analysis will focus on the differences in ERP data between the Greek and the Irish group of participants. With this analysis the second question (“Are the differences that are seen in Greek or Irish compared to the German similar to each other?”) is answered.

3. Expected results and discussion

Similar to the findings by Thierry et al. (2009), it is expected that there will be a visible reaction of colour change within the ERP data. This reaction will be present when the dark and light colour alternate in all blocks and does not require an active response of the participants. The visible reaction (called visual mismatch negativity (vMMN) in Thierry et al.) is expected to be similar for green and blue in German participants, thus not making a further distinction between different colour terms within blue or green.

This reaction will be in line with the findings for the English participant group in Thierry et al. For Greek participants, Thierry et al. established that there will be a different reaction for blue than for green, with a greater vMMN effect for the blue shapes. By adding an extra group, there is no known vMMN data for a language that distinguishes between terms for the colour green, such as Irish. However, it is to be expected that Irish participants will give a different response to green versus blue. It can also be expected that this difference in vMMN effect is the opposite of that of Greek participants. For Irish participants, it is expected that there will be a different reaction for green than for blue, with a greater vMMN effect for the green shapes. The assumption can be made that the greater vMMN effect for blues in Greek participants and the greater vMMN for greens in Irish participants will have similar characteristics, because having several terms for one colour category can be traced back to the same rationale, regardless of the colour itself.

The potential outcome from this study is that there is a difference in brain activity between languages with one term for each colour category (such as English, German, and Dutch) and languages with multiple terms for one colour category (such as Greek, Irish, and Russian). It is also expected that the brain activity in languages distinguishing two terms for blue (e.g. Greek) and the brain activity in languages distinguishing two terms for green (e.g. Irish) will share the same characteristics, and thus be similar.

Although the mentioned adjustments will give more insight into colour perception within different languages and show that the Saphir-Whorf hypothesis and the discussed literature could indeed be (partially) true, there is still a great deal of uncertainty about the influence of language on thought. In this paper it is assumed that the vMMN effect for the colour blue for Greek participants and the vMMN effect for the colour green for Irish participants share the same characteristics. This can be seen as a generalisation and future ERP studies should be carried out in order to prove or disprove this assertion. However, when proven correct, with the adjustments proposed it is still not clear what could be considered the origin of the differences in ERP data. For this reason, further research is needed to indicate a possible similarity or a possible difference in vMMN effects. Further research should also address the possibility that the differences in ERP data do not (solely) originate from the variance in colour terms in a language, and, if this is the case, where this difference might originate from. Another interesting change of perspective would be an alternative research design that focuses on multilinguals and the possible influence of one of their languages on their colour perception.

References

- Brown, R., & Lenneberg, E. (1954). A study in language and cognition. *The Journal of Abnormal and Social Psychology*, 49(3), 454-462.
- Coventry, K., Mitsakis, C., Davies, I., Jover, J., Androulaki, A., & Gómez-Pestaña, N. (2006). Basic colour terms in Modern Greek: Twelve terms including two blues. *Journal of Greek Linguistics*, 7(1), 3-47.
- Gallagher, S. (2017). *Enactivist Interventions*. Oxford, England: Oxford University Press.
- Han, Z., & Cadierno, T. (2010). *Linguistic Relativity in SLA. Multilingual Matters*. Berlin, Germany: De Gruyter.
- Heurley, L., Milhau, A., Chesnoy-Servanin, G., Ferrier, L., Brouillet, T., & Brouillet, D. (2012). Influence of Language on Colour Perception: A Simulationist Explanation. *Biolinguistics*, 6(4-4), 354-382.
- Kay, P., & Maffi, L. (2013). Green and Blue. In: M. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig, Germany: Max Planck Institute for Evolutionary Anthropology. Retrieved from <http://wals.info/chapter/134>.
- Koerner, E. (1992). The Sapir-Whorf Hypothesis: A Preliminary History and a Bibliographical Essay. *Journal of Linguistic Anthropology*, 2(2), 173-198.
- Regier, T., & Kay, P. (2009). Language, thought, and colour: Whorf was half right. *Trends in Cognitive Sciences*, 13(10), 439-446.
- Regier, T., & Xu, Y. (2017). The Sapir-Whorf hypothesis and inference under uncertainty. *WIREs Cognitive Science*, 8(6), 1-11.
- Simmons, D., Robertson, A., McKay, L., Toal, E., McAleer, P., & Pollick, F. (2009). Vision in autism spectrum disorders. *Vision research*, 49(22), 2705-2739.
- Swinkels, I. (2015). *Unusual features in the colour classification of Modern Irish* (Master's thesis). Retrieved from <https://openaccess.leidenuniv.nl/handle/1887/35015>.
- Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. (2009). Unconscious effects of language-specific terminology on preattentive colour perception. *PNAS*, 106(11), 4567-4570.
- Winawer, J., Witthoft, N., Frank, M., Wu, L., Wade, A., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *PNAS*, 104(19), 7780-7785.

Ted talking to his kids:

The Present Perfect in English and Dutch in ‘How I Met Your Mother’

F.J.G. (Veerle) Wilms

Research Master *Linguistics and Communication Sciences*, Radboud University Nijmegen

Abstract: Recent studies have tried to compare and understand the use of the present perfect (PP) in English and the *voltooid tegenwoordige tijd* (VTT) in Dutch (e.g., de Swart, 2007; Le Bruyn, van der Klis, & de Swart, 2019; Ritz, 2012; van der Klis, Le Bruyn, & de Swart, 2018). Most of these studies made use of literary translations to observe the difference between English and Dutch. A reoccurring statement in the existing literature is that the use of the PP and VTT is restricted in narrative discourse because English and Dutch do not allow for temporal relationships between eventualities. de Swart (2007), for example, states that it is not possible to tell a story in the PP or VTT. These conclusions were drawn on literary corpus research, however, instead of on spoken language. I propose that the use of the VTT in narrative discourse is broader than previously assumed, and that this can be observed from a study on ‘spoken language’ in the form of film script (lines from the series *How I Met Your Mother* (HIMYM)). A frequency count of the PPs and VTTs in two episodes of HIMYM revealed that the distribution of the VTT compared to the PP is bigger overall, but more importantly, it is bigger in narrative discourse. This shows that the differences between English and Dutch are not as clear-cut as existing literature makes it seem. The uses of the PP and VTT in spoken language have to be taken into account to fully understand the use of the PP and VTT.

Key words: present perfect, *voltooid tegenwoordige tijd*, Dutch, English, spoken language

1. Introduction

The present perfect is a construction consisting of an auxiliary verb and a past participle in both English (*I have eaten*) and in Dutch (*Ik heb gegeten*). The uses of this present perfect (PP) and *voltooid tegenwoordige tijd* (VTT) are different in the two languages despite their similar form. Recent studies have tried to compare and

understand the use of the PP and the VTT, and generally suggest that the VTT has a broader distribution than the PP (e.g., de Swart, 2007; Le Bruyn, van der Klis, & de Swart, 2019; Ritz, 2012; van der Klis, Le Bruyn, & de Swart, 2018). The current study adds to the understanding of the differences between the PP and the VTT by analysing the use of the two constructions in the comedy television series *How I Met Your Mother* (HIMYM). The findings of existing literature will be discussed first, before elaborating on the present study.

De Swart (2007) argued that the cross-linguistic differences of the PP across four languages are due to different discourse level properties instead of semantics. She compared the uses of the PP in four languages: French, German, English, and Dutch, in translations of the French novel *L'étranger*. The story of this novel was mostly written in *passé composé* (PC) rather than the “traditional literary tense” *passé simple* (PS) (de Swart, 2007, p. 2282). Translations of this novel can thus nicely show how the PC gets translated into German, English, and Dutch, and what the differences between these languages are regarding the PP. De Swart (2007) hypothesised that a temporal relation can be established between two eventualities with the French PC, but that this is not possible for the English PP and Dutch VTT. The PP and VTT should therefore not be as compatible with narrative discourse as the PC (also see Ritz, 2012, p. 20). It was furthermore hypothesised that the distribution of the PP is even more restricted than the VTT, because the PP does not “allow locating adverbials” (p. 2291). Her predictions were confirmed when analysing the translations of the novel, which can be observed from example (1) and (2).

1. a. Aujourd’hui, maman **est morte** (PC). Ou peut-être hier, je ne sais pas. J’**ai reçu** un télégramme de l’asile (PC): (...) p. 9
- b. Mother **died** today (SP). Or maybe yesterday, I don’t know. I **had** a telegram from the home (SP): (...) p. 9
- c. Vandaag **is** moeder **gestorven** (VTT). Of misschien gisteren, ik weet het niet. Ik **ontving** een telegram uit het gesticht (OVT): (...) p. 63
- d. Heute **ist** Mama **gestorben** (Perf). Vielleicht auch gestern, ich weiß nicht. I **habe** ein Telegramm vom Heim **bekommen** (Perf). p. 7
(de Swart, 2007, p. 2291)
2. a. Je me **suis réveillé** (PC) parce que j’avais (IMP) de plus en plus mal aux reins. p. 21
- b. I **woke** (SP) up because the pain in my back was getting (PPROG) worse. p. 17

- c. Ik **werd** (OVT) wakker omdat mijn lendenen hoe langer hoe meer pijn
begonnen (OVT) te doen. p. 71
- d. Ich **bin aufgewacht** (Perf), weil mein Kreuz immer mehr schmerzte (Prät).
p. 17
(de Swart, 2007, p. 2292)

All three occurrences of the PC in (1a) and (2a) are translated with a simple past (SP) in English, and two of the three occurrences are translated with an *onvoltooid verleden tijd* (OVT) in Dutch. This does not only show that the PC is used more in narration than the PP or the VTT, but these examples also show the slightly less restricted use of the VTT. The examples above present the opening lines of *L'étranger*, and the first PC is translated with a VTT in Dutch. VTT might not be fully compatible with narration, but it seems to be the case that VTT is acceptable at the start of narrative discourse. The use of the PP is more restricted than the VTT, because it is not preferred at the start of the story. The German translations of the PC in (1d) and (2d) are like a mirror image to the French sentences and show the *perfekt* (Perf), because, as de Swart (2007) explains, the German Perf "has basically the same discourse semantics as the French PC" (p. 2291). The author concluded that narrative use of the PP is possible in French and German because these languages allow for eventualities to enter temporal relations with other events, but that this is not so much the case for Dutch and English. The uses of the PP and VTT are therefore more restricted, as confirmed by the translations of *L'étranger*. Van der Klis et al. (2018) conducted a research which supports the claims made by de Swart (2007). They also analysed *L'étranger* and looked at the distributions of the verb forms which were used to translate the PC into Italian, German, Dutch, Spanish, English and Greek. The authors suggest that cross-linguistic variation is due to a competition between the perfect and the perfective past, and that French occupies an extreme position in this competition which results in a broad distribution of the PC. The PC is often translated with the Dutch OVT instead of VTT, and the English PP is hardly ever used to translate the PC. This pattern is the same as shown in de Swart's paper: The PC is often translated with OVT in Dutch, and even more so translated with an SP in English. Thus, the use of the VTT and PP are more restricted in narrative discourse. The conclusions about Italian, German, Spanish, and Greek are not discussed in this paper as these beyond the scope of the current study.

Another cross-linguistic analysis on the PP in English and Dutch has been done by Le Bruyn et al. (2019). The authors compared the English and Dutch versions of *Harry Potter and the Philosopher's Stone* (HP). This text contains both dialogue and narrative discourse, which is not so much the case in *L'étranger*. This allows for a comparison between the PP in dialogue and in narration. Their analysis of the novel revealed that

the PP and VTT are not used a lot in narration, and that the VTT is more broadly distributed than the PP in dialogue. Along similar lines as de Swart (2007) and van der Klis et al. (2018) proposed, Le Bruyn et al. (2019) conclude that the use of the perfect is “restricted to dialogue in both English and Dutch” (p. 5). Both *L'étranger* and HP seem to suggest that the VTT is not very compatible with narrative discourse.

The current study adds to this existing body of research by studying the use of the PP and VTT in spoken language instead of in literature. The following section elaborates on the research question and the hypotheses.

2. The present study

The three aforementioned studies have in common that they base their results on translated literary works. They did not find many instances of the VTT in narrative discourse, and it was therefore concluded that the use of the VTT is restricted in narration. This ‘restriction’ suggests that the use of VTT in narration is possible and grammatical nevertheless (as can also be observed from (1c)). Changing instances of OVT into VTT in example sentences from the previous literature reveals that this is indeed the case. (3a), (3b), (4a), and (5a) were taken from the existing literature. (4b) and (5b) are original translations for the current study.

3. a. *Ze **hadden** inderdaad een vrij gigantische hekel aan elkaar.* Past
they had indeed a fairly gigantic hatred at each other
“Well, they did rather detest each other.”
- b. *Ze **hebben** inderdaad een vrij gigantisch hekel aan elkaar **gehad**.* Perfect
they have indeed a fairly gigantic hatred at each other had
“Well, they did have hatred towards each other.”
- (Le Bruyn et al., 2019, p. 6)
4. a. (...) *ik weet het niet. Ik **ontving** een telegram uit het gesticht.* Past
I know it not. I received a telegram from the home
“I don’t know. I had a telegram from the home.”
- b. *Ik **heb** een telegram uit het gesticht **ontvangen**.* Perfect
I have a telegram from the home received
“I have received a telegram from the home.”
- (de Swart, 2007, p. 2291)

5. a. *Hij ging naar buiten, kwam weer terug en zette stoelen neer.* Past
 he went out, came back again and put the chairs down
 “He went in and out, arranging chairs.”
- b. *Hij is naar buiten gegaan, is terug gekomen en heeft de stoelen neergezet.* Perfect
 he is outside went, is back came and has the chairs put down
 “He went in, came back, and arranged the chairs.”

(van der Klis et al., 2018, p. 9)

In all sentences taken from *L'étranger* and HP above, a change from OVT to VTT does not lead to ungrammaticality. In fact, these are normal Dutch sentences. These plausible Dutch sentences do not appear in the analysed literature, however. It might be the case that translated literature does not provide sufficient information to draw definitive conclusions about the use of the PP and VTT in narrative discourse. Both *L'étranger* and HP are literary works, and the language used was most probably influenced by a literary style of writing. The existing observations on the PP and VTT are based on written literary language only, and not on spoken language. As spoken language could provide more insights into the uses of the perfect, spoken language has to be analysed as well before drawing conclusions about the behaviour of the VTT in the Dutch language. That is why this study compares the uses of the PP and VTT in narration and in dialogue in the American comedy television series HIMYM.

HIMYM is a sitcom about the fictive lives of five friends: Ted, Robin, Barney, Lily, and Marshall, who all live in Manhattan. Every episode of the show addresses everyday-life problems of these characters, mostly regarding themes as love, marriage, friendship, and career. This show was selected for analysis for several reasons. Firstly, HIMYM is an original series, so it has not been filmed and translated from a book. As a consequence, the series has been minimally influenced by traditional literary writing styles. Possible results on the VTT will thus not be affected by literary writing. Secondly, the show is framed in such a way that the main character Ted, in the year 2030, tells this son and daughter about the events that led him to meet their mother. He thus tells his kids stories about his life in past tense. The described events by Ted happened between 2005 and 2014, and are presented to a viewer of the show as if they are happening in present time. The show thus includes dialogue, i.e., when events play in chronological order, as well as narrative discourse, i.e., when Ted starts, interrupts, and ends an episode with his comments and stories in 2030. HIMYM provides an opportunity to analyse the uses of the VTT in narration and in dialogue, without this being a literary corpus study. Lastly, this television series provides a form of spoken language that is accessible and can therefore easily be analysed. HIMYM is an

American show, so viewers are presented with spoken English. Subtitles of this show can be downloaded from the internet for free. Subtitles are made by translators who transcribe what has been said in an episode, so Dutch subtitles are not simply a translation of the English subtitles. Dutch subtitles represent how a Dutch person would translate the spoken English. Because subtitles are such direct transcriptions of the spoken language, subtitles are used in the present study instead of, for example, the translated film-scripts of the series.

HIMYM thus offers an opportunity to analyse the uses of both the PP and VTT in spoken language. The instances of PP and VTT can be compared to each other, both in the narrative parts of the series and the parts with dialogue. The spoken language in a show like HIMYM can possibly give more insight into the uses of the PP and VTT than translated written language. Previous literature suggests that the VTT has a broader distribution than the PP, and that neither are often used in narrative discourse. This study aims to find out if these conclusions can also be drawn from spoken language. The present study addresses the question: Does the *voltooid tegenwoordige tijd* occur more in Dutch than the present perfect does in English in spoken narrative discourse in the television series *How I Met Your Mother*? Following the existing literature, it is expected that both the PP and VTT do not occur much in narrative discourse. Considering that the VTT in narration does not lead to ungrammaticalities, it is possible that the VTT can be found in spoken language more than it is found in literary works. The first hypothesis formulated regarding the research question is that VTT will be found in narrative discourse in HIMYM, more so than the English PP. Van der Klis et al. (2018) showed that the French PC is more often translated with a VTT in Dutch than with a PP in English. This indicates that the distribution of the VTT in narration is broader than the PP. The same is expected to be found in HIMYM. Secondly, previous research has suggested that the use of the PP is even more restricted than the VTT overall (de Swart, 2007; Le Bruyn et al., 2019; van der Klis et al., 2018). It is therefore also hypothesised that the VTT occurs more in dialogue than the PP.

3. Methods

3.1 Materials

Two episodes of the American comedy tv-series *How I Met Your Mother* were used for the analysis. This show in particular was used for this study, because besides dialogue it also contains narrative discourse in the form of the main character Ted telling his

future kids stories about his life. Episode 1 and 2 from season 7 of the show were selected because of availability reasons. The written subtitles in both English and Dutch were available for free on the internet, so these were downloaded. Subtitles were chosen for the analysis instead of, for example, an episode script, because the Dutch version of a script would likely be a translation of the English script. An analysis of a translated script is more like a translation of written language instead of spoken language. Secondly, scripts would provide a lot of information irrelevant to the current study, such as information about every movement of a character, background music, and for example eye-gazes. Only the spoken language and its direct translation to Dutch were necessary for this study, so subtitles were used.

3.2 Procedure

After the collection of the four files of subtitles (Episode 1 English, Episode 1 Dutch, Episode 2 English, Episode 2 Dutch), irrelevant information was deleted from the files. As subtitles also help hearing-impaired people while watching television, the subtitles would, for example, indicate whether music would play, or whether abstract sounds were made/played. This information was deleted, as it made the files messy and unclear. This left files which only contained the words spoken by the characters on the show.

Both episodes were watched twice while reading along with the subtitles; once for English and once for Dutch. This was done for two reasons: subtitles often contain mistakes, so it had to be made sure that no mistakes were made or words were omitted; and the subtitles did not indicate when normal dialogue was spoken or when Ted would provide information in the form of narrative discourse, so while watching the episodes these indications were added to the files manually.

The frequency counts for the use of PPs and VTTs happened in two stages. The first stage consisted of a search on a laptop, by searching for “have”, “has”, “s”, and “ve” for English and “ben”, “is”, “zijn”, “heb”, “hebt”, “heeft”, and “hebben” for Dutch with the search function (Ctrl + F) in the file. All occurrences of these verbs were checked to see if they were part of a PP or VTT. If they were, they were highlighted so they would be easy to find later. During the second stage, the printed versions of the files were checked for any forgotten or wrongly interpreted PPs or VTTs by reading the paper versions and correcting manually with a pen.

All PPs and VTTs were counted, and divided into two categories per episode: occurrence in narrative discourse versus occurrence in dialogue.

4. Results

The results of the search can be found in Table 1. As can be seen, only one English PP occurred in narrative discourse. This was not the case for Dutch: a total of five VTTs occurred in narrative discourse. Overall, the use of PP and VTT mostly occurred in dialogue.

Table 1

Number of present perfects / voltooid tegenwoordige tijden used in two episodes of How I Met Your Mother in English and Dutch.

Category	Episode 1	Episode 1	Episode 2	Episode 2
	English	Dutch	English	Dutch
Narrative discourse	0	2	1	3
Dialogue	12	27	8	37
Total	12	29	9	40

5. Discussion

This study aimed to find out if the Dutch VTT occurred in narrative discourse in spoken language. Two episodes of the sitcom HIMYM was analysed. It was furthermore investigated if the VTT would appear more in narration than the English PP. A frequency count revealed that a PP was used once in narration, while the VTT was used five times in total. This confirms hypothesis one, which stated that VTT would occur in narration in Dutch. The results also showed that the VTT was used more than the PP in dialogue, namely more than twice as many times in Episode 1 and over four times as many in Episode 2. This confirms hypothesis two, which predicted that the VTT has a broader distribution overall. This is in line with the findings of Le Bruyn et al. (2019).

These results show that the VTT is used less often in narration than in dialogue but that it is used nevertheless. Previous studies did not find many occurrences of VTTs in narration, and thus claimed that the use of the VTT is restricted in narrative discourse (e.g. de Swart, 2007; Le Bruyn, van der Klis, & de Swart, 2019; Ritz, 2012; van der Klis, Le Bruyn, & de Swart, 2018). It is possible that this restriction is greater in written language than it is in spoken language, because of, for example, formal writing styles. The VTT might not be suitable for a long story, but it is certainly acceptable to

use at the start or end of a conversation. Exactly this can be observed from HIMYM, as Ted uses the VTT to start or end his stories:

6. a. *En dat is hoe Marshall Punchy's bruiloft **verknoid heeft** (VTT).* Episode 1
and that is how Marshall Punchy's wedding ruined has
b. "And that is how Marshall destroyed (SP) Punchy's wedding."
7. a. *Jessica **is afgestudeerd** (VTT) op Rhode Island, en ze **heeft een** Episode 2
*wereldreis **gemaakt** (VTT).*
Jessica is graduated from Rhode Island, and she has a world
trip made
b. "Jessica is a Rhodes scholar. She's **travelled** (PP) all over the
world."*

In (6), Ted very explicitly ends his story about how Marshall ruined Punchy's wedding. The English version of the episode used an SP here, but the Dutch subtitle writer chose to translate this with a VTT. Example (7) shows how Ted starts the story about Jessica, in which a VTT is used twice. These two examples show that it is certainly possible to start or end a conversation using a VTT in Dutch in spoken language. The one occurrence of a PP in narration also happened at the start of a story, namely in the matching English sentence about Jessica, as can be seen in (7b). This possibility to introduce a new conversational topic with a perfect has been pointed out by Nishiyama and Koenig (2010). They argued that the perfect can be used for "topic negotiation" (p. 634). The authors conducted a corpus research and found that a perfect is sometimes used at the start of a conversation, to introduce a new topic. Ted also makes use of this strategy when he starts his story about Jessica. Note that this also happened in the Dutch translation of *L'étranger* (example (1c)). This indicates that not only the VTT can be used to start conversation, but a PP is also acceptable. This, however, only occurred once over both episodes, so no conclusions can be drawn about how commonly a PP is used like this.

It should briefly be noted that not much is known about the person or people who made the Dutch subtitles. It is most likely the case that these were made by one person, and not checked by other people. It is also unknown what the proficiency level in English is of this translator. The results are thus most probably based on the judgements of one Dutch person. It is plausible that a different translator would have used more or less VTTs. If a follow-up study would be conducted, it would be wise to have several translators write the Dutch subtitles. These different translations should then be compared for reliability. A second concern about the current study is that only two episodes of one show were analysed. Only one PP was found in these two

episodes, which might have been different if two other or more episodes were analysed. The same holds for the number of VTTs found. A future study in which more episodes, and ideally more series than just HIMYM, are analysed would provide a more complete picture of the distributions of the PP and VTT.

6. Conclusion

The present study showed that the Dutch VTT is compatible with narrative discourse, as it has been found in the translation of the show HIMYM in the short stories told by Ted. A frequency count revealed that VTT is used more often in narration than the PP, and that the VTT is used more than the PP overall. This second finding confirms previous findings that the VTT is more broadly distributed than the PP. Previous studies on the differences between the VTT and PP did not suggest that the VTT is compatible with narration. The current study differs from the existing literature in that it considered spoken language instead of written language. This study thus suggests that spoken language should be taken into account when trying to understand the uses of the VTT and PP.

References

- Camus, A. (1957). *L'étranger*. Paris, France: Gallimard.
- Camus, A. (1982). *The outsider* (J. Laredo, Trans.). London, England: Penguin Books. (Original work published 1957)
- Camus, A. (1994). *Der Fremde* (U. Aumüller, Trans.). Hamburg, Germany: Rowohlt Taschenbuch Verlag. (Original work published 1957)
- Camus, A. (1998). *De vreemdeling* (A. Morriën, Trans.). Amsterdam, The Netherlands: De Bezige Bij. (Original work published 1957)
- de Swart, H. (2007). A cross-linguistic analysis of the Perfect. *Journal of Pragmatics*, 39, 2273-2307. <https://doi.org/10.1016/j.pragma.2006.11.006>.
- How I met your mother. (n.d.). Retrieved from <http://www.tvsubtitles.net/tvshow-110-7.html>.
- Le Bruyn, B., van der Klis, M., & de Swart, H. (2019). The Perfect in dialogue: evidence from Dutch. *Linguistics in the Netherlands*, 36(1), 162-175. <https://doi.org/10.1075/avt.00030.bru>.
- Nishiyama, A., & Koenig, J. (2010). What is a perfect state? *Language*, 86, 611-646. DOI: <https://doi.org/10.1353/lan.2010.0014>.
- Radnor, J., Segel, J., Hannigan, A., Harris, N.P., Smulders, C., Bays, C., & Thomas, C. (Producers). (2006). *How I met your mother: The best man* [Television series]. Beverly Hills, Calif: 20th Century Fox Home Entertainment. Retrieved from Netflix.com
- Radnor, J., Segel, J., Hannigan, A., Harris, N.P., Smulders, C., Bays, C., & Thomas, C. (Producers). (2006). *How I met your mother: The naked truth* [Television series]. Beverly Hills, Calif: 20th Century Fox Home Entertainment. Retrieved from Netflix.com.
- Ritz, M. (2012). Perfect tense and aspect. In Robert I. Binnick (ed.), *The Oxford Handbook of Tense and Aspect*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195381979.013.0031>.
- van der Klis, M., Le Bruyn, B., & de Swart, H. (2018). *A multilingual corpus study of the competition between past and perfect in narrative discourse* (Unpublished Manuscript). Utrecht University, Utrecht.

