

What makes a catchphrase catchy?
Possible Determinants in the Borrowability of English Catchphrases in Dutch
Eline Zenner, Dirk Speelman & Dirk Geeraerts

1. Introduction

This paper presents a multivariate corpus-based analysis of the spontaneous use of English catchphrases in Dutch. Taking a cognitive sociolinguistic perspective, we show how an interplay of socio-conceptual, encyclopaedic and linguistic characteristics of the catchphrases under scrutiny can help explain whether a given catchphrase occurs in Dutch newspapers.

As such, this paper serves as an exploratory case-study for a broader research project in which we try to overcome two shortcomings in existing anglicism research. Firstly, by incorporating a variety of lectal, pragmatic and socio-conceptual features in the analysis, we wish to widen the theoretical scope of the predominantly structuralist paradigm of anglicism research. Secondly, this usage-based study serves as empirical support for theoretical claims that have recently been made concerning the need for analyses of borrowing (English) formulaic sequences.

Although catchphrases form a niche of formulaic sequences, it is made clear that they form a particularly useful and interesting starting point for this project, as they can help shed light on the importance of mass media in the spread of English. We also indicate how this is especially meaningful for research in language communities where English is primarily used for international communication.

The structure of this paper is as follows: in a first section, we develop the theoretical claims we wish to make in this paper and the way we have translated these into the presented case-study on catchphrases. Secondly, we present the design of our study: after introducing the data, we discuss the selection of dependent and independent variables. Special attention is paid to both advantages and drawbacks of the chosen predictors, when compared to alternative operationalizations. In the next section, the results of the study are presented. We discuss the output of a logistic regression analysis, a multivariate technique which takes the interplay of all variables into account. After presenting the main model, we highlight whether any lectal variation is seen in the importance of the different predictors. Conclusions are made in the final section, where we also mention future steps we wish to take.

2. Theoretical Background & Research Question

The aim of this project is to overcome two shortcomings in anglicism research. Firstly, a lack of attention for socio-pragmatic and cognitive aspects is to be noted in research on English loanwords. Secondly, neither the contact linguistic paradigm nor the phraseology framework have as yet inquired into the mechanisms underlying the borrowing of formulaic sequences. We now briefly discuss each of these shortcomings in more detail. Next, we introduce how our case-study, a multivariate analysis of the use of English catchphrases in Dutch, serves as a first step in tackling both issues. We also stress the theoretical and practical reasons that led us to take this subset of formulaic sequences as the starting point for the project.

2.1 Borrowability of formulaic sequences: a cognitive sociolinguistic perspective

2.1.1 A socio-cognitive Perspective on Borrowability

Our analysis focuses on the use of English in Dutch, as it is spoken in Flanders and the Netherlands. Both regions belong to Kachru's "expanding circle", a cover term used for all language communities where English does not have any official status, but mainly functions as language for international communication (see Kachru 1992, but also e.g. Yano 2009 for criticism on the model). Within this expanding circle, and more specifically within German linguistics, a longstanding tradition of anglicism research has emerged since the post-war period (e.g. Carstensen 1965). Over the years, the "corpus-based" side of this tradition has developed and maintained a narrow focus on structuralist analyses, both in general as with respect to the issue of borrowability.

In general, the main tasks anglicism research has given itself are the counting and classification of anglicisms. As a first step in this type of research, a sample of anglicisms is collected. These samples are often based on *ad-hoc* personal selections of the researchers, for which it is unclear from what type of data they are drawn and hence how representative they really are (e.g. Zandvoort 1964, Kurth 1998). A more methodologically sound way of sampling is seen when researchers create databases of anglicisms based on actual corpora. Unfortunately, these corpora are always limited in size, as the anglicisms are extracted manually (e.g. Fink 1997, Onysko 2007). The second step is then to group the anglicisms in the samples according to several dimensions, the most popular of which is the level of adaptation to the source language. The main focus is on morphological adaptation, but attention is also given to phonetic, orthographic and syntactic integration (e.g. Filipovic 1977, Nettmann-Multanowska 2003). Other types of classification are for instance based on the lexical field the items belongs to (e.g. Krauss 1958, Grigg 1997).

When dealing with the issue of borrowability, which focuses on the identification of factors that have an influence on the ease with which items can be borrowed, we see a similar structural focus emerging. Aside from the exploration of influential factors like genealogical relatedness of languages (see Haspelmath 2008), the main tradition in this field is to create clines of borrowability based on parts of speech (Whitney 1881, elaborated on by e.g. Haugen 1950 and Field 2002: 117). In anglicism research, the validity of these clines is often verified by means of basic counts of types per word class (e.g. Yang 1990: 29, Onysko 2007: 131). The problematic aspects of these clines, and most importantly the question of what factors can explain the higher borrowability of certain parts of speech, are rarely pointed out (see Van Hout & Muysken 1994 for a description of the issue). Furthermore, other possibly influential features for the borrowability of anglicisms are left unmentioned.

Although presented from a bird's eye view, the structuralist nature of the predominant strand in anglicism research should be clear. Nevertheless, it is important to nuance this view as, over the years, occasional attempts have been made at incorporating a more social perspective on the issue of (borrowability of) anglicisms.

First of all, in their study on the use of English loanwords in five neighbourhoods in Canada, Poplack *et al.* (1988) link the degree of integration of anglicisms to several social factors. Using a large database of anglicisms and employing multivariate statistical analyses, they show the importance of social class, neighbourhood, age and

sex for explaining variation in the level of integration of anglicisms. Although this approach is promising, two problems remain. Firstly, the research question on integration is still predominantly structuralist. Secondly, the selected social features are typical of first-wave sociolinguistics, of which several shortcomings have been mentioned over the years (Eckert *under review*).

Secondly, Gerritsen *et al.* (2007) broaden the field of anglicism research to the discourse level and incorporate a lectal approach. In their study on the use of English in advertisements in six countries, it is shown that the use of English is typical for distinct parts of an ad. Furthermore, attention to lectal variation comes up in the comparison of the use of English in Belgian Dutch and Netherlandic Dutch (no difference between the regions is found). This incorporation of lectal variation is innovative, but limited: a broader variety of socio-lectal features needs to be studied.

Finally, Hilgendorf's (2007) plea for a sociolinguistic approach to English influence is to be situated in the macro-domain. In discussing the broad functions of English and English intrusion in several communicative domains (e.g. politics and law), her position moves away from quantitative anglicism research towards the more qualitative English Lingua Franca-research (e.g. Seidlhofer 2009). This macro-sociolinguistic approach makes her study less useful for our purposes.

Overall, it is safe to say that, despite these attempts, anglicism research is in need of a cognitive sociolinguistic approach, as it was developed in (amongst others) Geeraerts (2005) and Kristiansen & Dirven (2008). In view of the narrow focus of borrowability research on formal linguistic features, this paper pleads for the incorporation of (1) socio-lectal features, without limiting these to typical first-wave predictors (e.g. Ruelle *et al.* 2010; Zenner *et al.* 2009a); (2) encyclopaedic and cognitive/conceptual features, tailored to the phenomenon under scrutiny (see Geeraerts *et al.* 1994); (3) pragmatic and stylistic features (e.g. De Sutter *et al.* 2008) and (4) the combined effect of all of the above, by using multivariate statistical analyses (e.g. Impe *et al.* 2008, Levshina *et al.* 2010).

2.1.2 Borrowed Formulaic Sequences

Aside from the structural nature of anglicism research, its narrow focus on loanwords and compounds is remarkable; the possibility of borrowing formulaic sequences (e.g. *as good as it gets*, *oh my God* etc.) has largely been ignored. With only very few exceptions, research on the use of longer stretches of foreign language material has been limited to the domain of code-switching (e.g. Muysken 2000, Deuchar *et al.* 2007). However, it is not hard to appreciate that the borrowing of formulaic sequences and the actual mixing of codes are two very different phenomena.

First of all, code-switching research deals with the spontaneous mixing of *ad hoc* created material from different codes (example 2.1, drawn from Poplack 1980). The use of foreign formulaic sequences, on the other hand, does not involve this creation of new language material; the expression can be retrieved as a whole (example 2.2, taken from our Dutch newspaper corpus).

(2.1) And from there I went to live PA' MUCHOS SITIOS. Después viví en la ciento diecisiete WITH MY HUSBAND. (Poplack 1980: 597)

(2.2) Mijn favoriete beginsel is dat ieder mens recht heeft op the pursuit of happiness. (De Morgen, 21/10/2000)

Secondly and related to this, there is a clear difference in the proficiency level required for the phenomena. As Poplack (1980) points out, code-switching is typical for bilinguals who are highly proficient in both codes. By contrast, bilingualism is no prerequisite for the use of foreign formulaic sequences at all.

Despite the lack of attention for borrowed sequences, which is also present in the field of phraseology, there has always been “the generally recognized possibility of borrowing idiomatic phrases as units” (Pfaff 1979). Moreover, the rise of English as a language of international communication (e.g. Crystal 2003) makes an increase in the borrowability of English formulaic sequences in the “expanding circle” very plausible. This has not gone unnoticed: over the last few years, attention for the phenomenon has started to creep up in anglicism research.

For instance, Sharp (2001) pays attention to longer stretches of English in spoken Swedish. However, she uses the term *code-switching* as an aggregate for all English multi-word units in her corpus, although most of these classify as borrowed formulaic sequences (e.g. *go to hell, be my guest*). She hints at the formulaic nature of the phrases herself (Sharp 2001: 107), but does not link this to a theoretical distinction between code-mixing and borrowed phrases. A similar problem appears in Onysko’s work (2007), which focuses on written German. He dedicates a paragraph to “multi-word phrasal borrowings” (e.g. *state of the art, just in time*) but fails to be consistent in distinguishing these from his examples of intra- and inter-sentential code-switching (e.g. *the place to be*). A more promising step has been taken by Androutsopoulos (*in press*). In his theoretical paper, he clearly states the “lack of a meso level framework in the study of English in the mediascapes of the ‘expanding circle’”. Some interesting theoretical points are raised concerning the textual functions of this ‘English on top’. However, only preliminary examples are provided to support his theoretical claims.

2.1.3 A Socio-cognitive Perspective on the Borrowability of Formulaic Sequences

Above, we have shown how the scope of anglicism research needs to be expanded to the socio-pragmatic and cognitive field. Secondly, we indicated the need for a framework on the borrowability of formulaic sequences. Our project brings both aspects together by raising the following question: from a group of socio-conceptual, linguistic and encyclopaedic features, which are influential in determining the borrowability of formulaic sequences? In the current paper, we take the first steps in resolving this issue by investigating the usability of English catchphrases in Dutch.

2.2 Case-Study: Spontaneous use of English catchphrases in Dutch

Catchphrases are expressions used in (visual) media, politics, literature etc. that ‘catch on’ and get incorporated in “the phraseological component of the native speaker’s lexicon” (Alexander 1983: 11). Hence, they are used freely in discourse, in contexts detached from the original source. Henceforward, we refer to this non-source related use of a catchphrase as ‘spontaneous use’.

Based on this definition and on the background sketched above, the main question for this case-study is what factors influence whether an English catchphrase occurs spontaneously in Dutch, and whether these mechanisms are consistent across lects. Before dealing with the actual methods we developed, we briefly discuss why

catchphrases form the starting-point for this research. Although they are a rather unprototypical subset of formulaic language (see Moon 1998: 22), catchphrases are interesting from both a practical and a theoretical point of view.

Firstly, Van Hout & Muysken (1994) state that issues of borrowability are best discussed by means of set-external proof, in the form of a source-language corpus. The point is to have both positive and negative evidence; verifying which set of elements from a source corpus end up in the recipient language and which do not, is a reliable way to identify factors that influence borrowability. For our project, this means we should create a list of English formulaic sequences. We can then check which elements from this list occur in Dutch. In this respect, catchphrases are particularly useful, as quite some people find pleasure in posting lists of English catchphrases online.

Secondly and more importantly, the media-related origin of catchphrases makes them interesting from a theoretical point of view. Intensity and source of language contact have often been mentioned as crucial for borrowability (see Haspelmath 2008). More specifically for the expanding circle, the asymmetrical contact with English via mass media is often mentioned as the principal factor in its spread (see Kowner & Rosenhouse 2008, Androutsopoulos *in press*). However, due to the methodological complexity of the phenomenon, no empirical evidence for this presumed importance of media-influence has been given so far. Moreover, research on monolingual language use minimises the role of mass media in language change (see Labov 2001: 356, 363-364). Hence, the actual contribution of media influence to the rise of English still is an unresolved issue.

The occurrence of English catchphrases in Dutch can serve as a first step in dealing with this issue: as English movies and series are subtitled instead of dubbed in Flanders and the Netherlands, the importance of the original source for the borrowability of the catchphrase can be verified. As such, our main research question becomes:

how important is (the entrenchment of) the media source of the catchphrase for its borrowability, compared to the influence of other socio-conceptual, linguistic and lectal features?

Of course, this specific comparison of features can only serve as a first step in determining the role of mass media. Specifically, our question focuses on the comparison of the importance of media with other features in the spread of a media-related phenomenon. Obviously, this research needs to be complemented with studies that compare the importance of media in the spread of media-related expressions like catchphrases on the one hand and of other types of fixed expressions (e.g. figurative idioms) on the other hand.

3. Data, Variables & Method

The main method for this research is roughly comparable to the analysis performed by Van Hout & Muysken (1994) and consists of four main steps. First, we create an external list of English catchphrases. Second, we verify which of these catchphrases occur (spontaneously) in a large Dutch newspaper corpus. Third, we identify a set of predictors that might explain the success or failure of a catchphrase. For this step, our cognitive sociolinguistic approach is crucially different to Van Hout & Muysken

(1994), who only incorporate formal linguistic features. Finally, the importance of each of our predictors is determined by performing multivariate statistical analyses. The specifics for each of these steps are discussed below.

3.1 Data

3.1.1 External catchphrase list

As noted above, the practical advantage of working with catchphrases is the existence of online catchphrase lists. We collected all catchphrases from nine such lists, which resulted in a first set of more than 1000 types.

To exclude personal preferences of the list makers (who sometimes show a disproportionate and unrepresentative fondness for certain movies or series), only those catchphrases that occur twice in our original set, are considered for the analyses. Two exceptions are made. Firstly, one of the nine lists has been created by an American cable television network (TV Land), who used the list in the TV special “The 100 Greatest TV Quotes & Catchphrases”. Because of the wider scope of this list, all 100 TV Land catchphrases are incorporated in the dataset. Secondly, to avoid underrepresentation of UK sources, all catchphrases with a UK origin (see 3.1) were selected. Nevertheless, our database still has a rather unequal distribution of both regions. We will come back to this below.

- (3.1) Nudge, nudge, wink, wink, say no more (Eric Idle in *Monty Python’s Flying Circus*)

As our research question focuses on the spread of media-related catchphrases, a further restriction of the final list is required: we only include catchphrases from movies (see example 3.2), series (example 3.3) and TV shows (example 3.4). All other types of catchphrases (example 3.5) are removed.

- (3.2) You can’t handle the truth! (Colonel Nathan Jessup in *A Few Good Men*)
(3.3) The truth is out there. (Fox Mulder in *The X-Files*)
(3.4) The tribe has spoken. (Jeff Probst in *Survivor*)
(3.5) I did not have sexual relations with that woman. (Bill Clinton)

The eventual list contains 229 catchphrases. Of course, the fairly small size of this set is a direct consequence of the restrictions. An alternative approach would be to incorporate all the catchphrases we find in the nine online lists. However, applying the strict selection criteria we have chosen here, forms the most appropriate way to get a first indication of the borrowability of catchphrases.

3.1.2 Corpora

The next step in the analyses is to verify which catchphrases occur (spontaneously) in Dutch. This step relies on two large newspaper corpora, which represent the two main national varieties of Dutch. LeNC, the corpus for Belgian Dutch, consists of data from six different newspapers from 1999 to 2005 and contains over one billion words. TwNC forms the Netherlandic Dutch counterpart for LeNC: it contains roughly 300 million words, for five different newspapers in the period from 1999 to 2002. For the main analysis both corpora are taken together. In a next step, we verify whether any

lectal variation exists. We specifically inquire into regional variation by comparing the occurrence of catchphrases in TwNC and in (the correct subset of) LeNC.

Although we do not expect newspapers to be swamped with catchphrases, we have some convincing arguments to base the analyses on this type of language use. Firstly, it allows us to stay true to the tradition of using newspaper corpora in anglicism research. At the same time, we are able to move away from the tradition by using a corpus that is sufficiently large to make reliable claims on low frequent phenomena like catchphrases. So far, anglicism research has not used corpora of over five million words, due to the manual extraction methods we mentioned before. Secondly and more importantly, newspaper corpora form a stable and reliable source to determine whether a given item has penetrated the language at large. As newspaper journalists write for a big audience, they do not easily use expressions they deem unknown to their readers. Hence, any catchphrase used (spontaneously) in newspapers is most likely familiar to the average speaker of Dutch. We will come back to the verification of this claim later.

3.2 Response: spontaneous use of English catchphrases in the Dutch corpus

The identification of the dependent variable of this study relies on determining which of the 299 catchphrases occur in our newspaper corpus. Using a broad search, we are able to find all potential hits, including shortened (example 3.6) and slightly altered (example 3.7) versions of the catchphrases.

- (3.6) Het wordt steeds duidelijker dat deze Russische peetvader Poetin wel eens echt ten val zou kunnen brengen . Maar niet met “an offer he can’t refuse”. (DM 08/12/2005)
- (3.7) Dezelfde krant kon in elk geval de verleiding niet weerstaan om te koppen “Don't mention the score”, met een knipoog naar de bekendste Fawlty Towers-aflevering. (NRC 12/12/2001)

After deleting all noise from the list, 1598 observations for 96 catchphrases remain. A next and very crucial step is now to distinguish three different ways in which a catchphrase can manifest itself in our corpus and to appreciate that only one of these shows the actual intrusion of English catchphrases in the Dutch phraseological lexicon.

Firstly, the SOURCE-RELATED USE of the catchphrase is found when the catchphrase occurs in a piece of text that deals with the movie or series it originates from (see 3.8). These 656 observations are no indication of the actual penetration of the catchphrase in the Dutch phraseological repository: they only serve as an indication of the entrenchment of the source of the catchphrase in Dutch media. What is measured here is how often journalists for example talk about James Bond, not how often they freely use the expression “shaken not stirred” in discourse.

- (3.8) Sean Connery gaf toen gestalte aan de Britse geheime agent die zijn wodka “shaken, not stirred” drinkt en tussen de vuurgevechten door 's werelds mooiste vrouwen in bed praat. (LN 09/04/1999)

Secondly, some catchphrases are used to NAME companies, events or organizations. A Belgian record company for instance chose to name itself “Play it Again Sam”, and an artistic photographer used “there’s no place like home” as the name for his exhibition. All 412 of these observations are excluded from the analyses.

Finally, the SPONTANEOUS USE of catchphrases is seen when the catchphrase is used in texts on subjects that are not related to the movie or series the catchphrase belongs to (see examples 3.6 and 3.7). Only these 530 observations reflect the actual intrusion of the catchphrase in Dutch. Hence, only this use of the catchphrase is regarded as the dependent variable for this study.

As a final remark on the definition of the dependent variable, we highlight the possibility of making a further subdivision within the category of “spontaneous use” between situations where the origin of the catchphrase is still explicitly mentioned (example 3.7) and situations where the catchphrase is used completely detached from its origin (example 3.6). We have aggregated over this distinction for our analyses, but it will be developed in future research.

3.3 Predictors: Possible determinants in the borrowability of catchphrases

Once the response variable has been identified, the factors that may influence its behaviour have to be defined. For the study of spontaneously used English catchphrases, we compare media influence with the entrenchment of the catchphrase in English and with its formal and pragmatic features. Below, we discuss the operationalization of these groups of features. We also present possible alternatives and indicate what restraints led us to implement the predictors as presented here.

3.3.1 Direct Media Exposure: Entrenchment of the source of the catchphrase

Our first feature, DIRECT MEDIA EXPOSURE, is used to indicate how familiar a Dutch language user is with the source of the catchphrase. Put differently, it measures the importance of the popularity of the movies or series in which the catchphrase was first used for its borrowability. We operationalize this feature in two ways.

First, we check the Google frequencies for the title of the source of the catchphrase, restricting the search to pages in Dutch. We complement the query with the words “film” or “serie” (e.g. for “beam me up, Scotty”, the query would be “‘Star Trek’ + serie”). The results are classified in 3 frequency bands (3.9).

- (3.9) frequency bands:
- (a) less than 5000
 - (b) [5000 – 25 000[
 - (c) more than 25 000

Although Google frequencies are a practical and efficient operationalization of media entrenchment, they form a rather indirect and oblique measure of the popularity of the source. A more tangible alternative would be to work with box office figures and audience ratings. We pursued this possibility, but were hindered by the unhelpfulness of one of the most important commercial broadcasting corporations in Belgium.

Second, we also used the occurrence of SOURCE-RELATED USE of a catchphrase in our corpus (see 3.8) as an indication for the popularity of the source. We use a binary classification (occurrence or absence of source-related use), which is more practical than working with raw frequencies. We chose to work with the occurrence of source-related use (e.g. “beam me up, Scotty”) over the occurrence of the title of the source (e.g. *Star Trek*) in the newspaper corpus, to ensure that the two operationalizations for media entrenchment measure different phenomena. This way, we can provide a more nuanced view on the issue.

3.3.2 Nature of Media Exposure: Encyclopaedic characteristics of the source

With the next group of features we focus on the encyclopaedic characteristics of the source of the catchphrase and on what these may tell us about the nature of the exposure to English media in Flanders and the Netherlands. All information is based on Wikipedia and the Internet Movie Database. A first variable is the COUNTRY OF ORIGIN of the source of the predictor. All selected movies and series either come from the UK or the US. This predictor can help verify whether the United States are, as is often claimed, the primary locus of English influence. Secondly, we compare SOURCE-TYPES, by verifying whether movies generate more spontaneous catchphrases than series or shows. Finally, we determine the importance of the AGE OF THE SOURCE, by comparing catchphrases younger than forty to catchphrase older than forty. For movies, age is based on the release date in the country of origin. For series, it is based on the date of the first broadcasting of the first episode, also in the country of origin.

Of course there are other encyclopaedic features to think of. However, most of these are only applicable to either movies or series. We for instance tested the influence of the genre of the source, but as comedies are highly overrepresented for series, the predictor is only useful for the analysis of catchphrases from movies.

3.3.3 General Exposure to the Expression: Frequency of the catchphrase in English

The following feature is meant to capture the POPULARITY OF THE CATCHPHRASE IN (THE INTERNATIONAL USE OF) ENGLISH. The goal is to determine to what degree Dutch users are exposed to the catchphrase, aside from their contact with its original occurrence in the media. This way, we capture the possibility that Dutch users might be familiar with the catchphrase without having had any contact with the original media source. We operationalize this factor by checking the Google frequencies of the catchphrase, limiting the search to pages in English (e.g. 213 000 hits for “beam me up, Scotty”). We then classified the results in three frequency bands:

- (3.10) (a) less than 100 000
- (b) [100 000 – 1000 000[
- (c) more than 1000 000

By searching all Internet pages written in English, this approach uses the broadest possible definition of exposure to English, which is ideal for exploratory analysis. However, it obscures certain divisions between variants and varieties of English. Specifically, “pages in English” covers a load that is most likely broader than the varieties of English that Dutch speakers stand in contact with. Hence it is advisable to get a more disentangled view in future research. The most straightforward approach would be to focus on the use of the expression in corpora for ELF and for the two “traditional” varieties of English (UK and US).

3.3.4 Linguistic Features of the Catchphrase

The formal and pragmatic characteristics of the catchphrase form the final group of potentially influential features we will identify in this study. The first of these predictors is the NUMBER OF WORDS in the catchphrase. We base this on the catchphrase as it is found in the online lists, not as it is used in the Dutch corpus. The second variable indicates the percentage of NON-CONVENTIONAL VOCABULARY in the

catchphrase. Exclamations, onomatopoeia and proper names are regarded as non-conventional. The percentage is computed by taking the ratio of the amount of non-conventional words to the total number of words (see examples 3.11 to 3.13).

- (3.11) Yippee-ki-yay, motherfucker! (John McClane in Die Hard; ratio = 50%)
- (3.12) D'Oh! (Homer Simpson in The Simpson; ratio = 100%)
- (3.13) Use the force, Luke. (Obi-Wan Kenobi in Star Wars; ratio = 25%)

The COMMUNICATIVE FUNCTION of the catchphrase is the last of our predictors. We distinguish three groups, based on sentence types: (1) statements (see 3.14), (2) interrogatives and requests (see 3.15), and (3) exclamations and commands (see 3.16)

- (3.14) I know nothing, I'm from Barcelona (Manuel in Fawlty Towers)
- (3.15) Permission to speak, Sir. (Jones in Dad's Army)
- (3.16) Alllll-righty, then! (Ace Ventura in Ace Ventura)

Existing functional models of formulaic sequences in English (e.g. Lattey 1986, Moon 1998: 215 ff. and Wray & Perkins 2000) are less applicable to our analyses. As we are dealing with the borrowability of catchphrases, we have to assign pragmatic function based on types, not on tokens. Hence, our operationalization is inevitably basic. Future analyses of the corpus examples will allow us to incorporate more advanced stylistic and pragmatic analyses.

3.4 Analyses

A first glance at the predictor shows that 58 of the 229 inspected catchphrases occur spontaneously in our corpus. The final step in our study is then to determine why it is precisely these 58 catchphrases we find. More specifically, we want to identify which of our independent variables have a true influence on the borrowability of English catchphrases in Dutch. To answer this question, we perform forward stepwise logistic regression analyses. This multivariate statistical technique determines which independent variables have an influence on the behaviour of a binomial dependent variable, taking the combined effect of all the predictors into account. Performing the analysis "stepwise" ensures that the variable with the strongest impact will be selected first, followed by the second most important factor and so on.

For our analyses, this means we verify which predictors influence *whether* a catchphrase occurs, *not how often* it occurs. The 229 types form our set of observations, the response is binary ("does or does not occur spontaneously in the corpus") and the specific token frequency of the 58 occurring types is disregarded. Although disregarding the frequencies might appear to be a limited approach, there are several good reasons to employ this technique. First of all, there is not much variation in the token frequencies of the occurring catchphrases: 23 occur only once, the median of the frequencies is 3 and the maximal token frequency is 71. Taking the overall size of the corpus into account, it is clear that the token frequencies lie closely together. Moreover, the frequencies are not normally distributed, which makes other multivariate techniques inapplicable. Secondly, Van Hout & Muysken (1994) also choose forward stepwise logistic regression analysis: when determining influential predictors in borrowability, the sheer occurrence of an expression or lexeme is by far more important than the number of times it occurs.

Before we present the results of the regression analysis, one final comment is required on the status of single occurrences in this research. As mentioned above, 23 of the 58 spontaneous catchphrases occur only once. Taking the size of our corpus into account, the question begs itself whether these single occurrences can tell us something about language at large, or whether they are to be regarded as whimsical and unreliable. For our specific dataset, we have reasons to believe we can form conclusions based on all 58 catchphrases, including the single occurrences. First of all, we have to take the nature of our corpus into account. As mentioned before, newspaper language is stable and traditional. Journalists take their large audience into account and will not be prone to use unknown or uncommon constructions. Hence, even a single occurrence of a catchphrase can be seen as a clear indication of its overall use. Secondly, we have supported this claim by performing our regression analysis two times: once for the model with single occurrences, once for a model that only incorporates catchphrases with two or more spontaneous occurrences in the corpus. Without going into the details of these regressions, it is safe to say they corroborate our claim: the model with single occurrences did not overfit the data. Hence, the single occurrences can be discussed in the analysis.

4. Results

For the discussion of the regression output, we first focus on the main model, which takes the results of both corpora together. Next, we examine possible lectal variation in the importance of the predictors, by comparing the regression models for the Belgian Dutch and the Netherlandic Dutch corpus.

Before discussing the results, a final comment has to be made on how to read the output of the model. For each variable, we take one value as reference point (e.g. “declarative” for sentence-type). Then, the analysis verifies which of the other values (e.g. “question/request”) has an effect on the success-rate of the response variable, compared to that of the reference point. Only variables with significant effects ($p < .05$) are incorporated in the model. The reference point itself is never included. The sign of the estimate shows the direction of the effect: if this is a positive number, it means we find a higher success-rate for the given value, compared to the reference point. For this study, a higher success-rate means that more catchphrases will be used spontaneously in the corpus. Negative estimates imply a lower amount of spontaneous catchphrases.

4.1 Main Model

Our discussion of the main model (Table 4.1) starts off with an overview of the general characteristics of the regression output. We then discuss the significant predictors, to round off with some remarks on the insignificant predictors.

4.1.1 General characteristics of the model

The general diagnostics for logistic regression confirm the goodness of fit of the model. The chi-square test is significant and the data do not exhibit overdispersion. Also, no multicollinearity exists, which means that there are no inordinately strong

correlations between the independent variables in the model. It is thus safe to say that our model contributes to the explanation of the borrowability of catchphrases in Dutch.

An important question is then how big this contribution is. We use two measures for this, which both show that our model is quite powerful. Firstly, R^2 is used to indicate how much of the variation can be explained by the model. In our case, the R^2 is a very acceptable 52.6%. Secondly, the C-measure is used to verify whether a given model has any predictive power. Our C-level of .89 means that, given new sets of catchphrases, this model will, to a reasonable extent, succeed in predicting which ones occur spontaneously in Dutch.

Coefficient	Estimate	Standard Error	z-value	p-value
(Intercept)	-1.47	0.46	-3.23	0.001
SRU: yes	2.67	0.43	6.27	0.000
exposure in English (Google): 1	0.92	0.28	3.33	0.001
exposure in English (Google): 2	0.33	0.15	2.16	0.031
sentence: question/request	0.18	0.62	0.29	0.773
sentence: exclamation/command	-1.40	0.46	-3.02	0.003
type: series/show	-0.83	0.42	-1.99	0.047

Table 4.1

4.1.1 Significant predictors: entrenchment in Dutch and exposure to English

Now that we have established the overall significance and importance of the main model, we can focus on the actual effects we find. Following the stepwise technique, we present the significant predictors in order of importance.

The occurrence of the SOURCE-RELATED USE of the catchphrase in our newspaper corpus, which measures DIRECT EXPOSURE TO THE MEDIA SOURCE, is the most important factor in the model. We can deduce from Table 4.1 that catchphrases which also occur in contexts related to the source of the catchphrase (see 3.8) will more often have spontaneous occurrences in the corpus (see 3.7) than catchphrases without source-related use. This entails that the familiarity Dutch speakers have with the original origin of the catchphrase has a significant effect on the borrowability of the catchphrase. However, final conclusions on the importance of direct media exposure can only be made once we compare this factor to the importance of the popularity of the catchphrase in (International) English.

Indeed, Table 4.1 shows that the OVERALL EXPOSURE Dutch users may have to the expression in (INTERNATIONAL) ENGLISH is also an important predictor in the model. Taking the ordinality of the Google frequency measure into account (using Helmert coding), we find a steady rise of the amount of spontaneously used catchphrases from the lowest to the highest frequency band. This means that the more popular a catchphrase is in English, the bigger the chance it is used spontaneously in our corpus.

Bringing both variables together, we see that (1) direct media exposure and (2) indirect exposure to the expression via (International) English are both important features for the spread of catchphrases. More specifically, the results presented above indicate that a catchphrase is likely to occur spontaneously in the corpus if it is used in Dutch in relation to its media source and if a large group of English speakers uses the

expression. Of course, it is important to verify whether both claims are independent of each other and if not, what this entails for the interpretation of the predictors. If a large (positive) correlation coefficient is found, this would mean that both variables are measuring something very similar. Hence, asking the question if Dutch speakers are familiar with the catchphrase through the original media source or through exposure to the expression in their contact with English, would become hazardous.

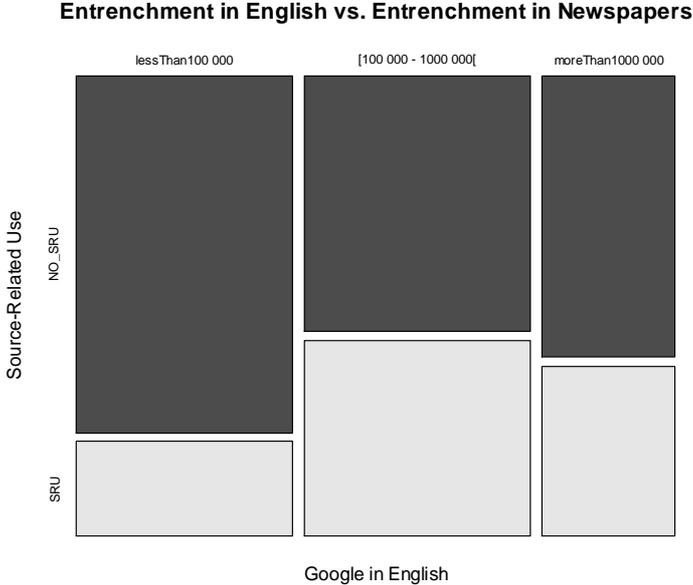


Figure 4.1

As is visible in Figure 4.1, we find a correlation between both features (p for spearman = .01, rho = .17). Specifically, we see that catchphrases that are located in a higher frequency band (which means that they are used more often by English speakers) have a higher chance of having source-related uses (SRU) in our Dutch corpus. Although the correlation is too weak to jeopardise the conclusions we made above, it is important to interpret it. The main question is for what reason the popularity in International English and the media-related occurrence of a catchphrase in Dutch newspapers are linked. Is the entrenchment in English also influenced by the popularity of the series or movie the catchphrases originates from, or is the SRU in Dutch influenced by the popularity of the expression in English?

Crucially now, disentangling this chicken-or-egg problem is complicated by the problems we mentioned above concerning the use of Google frequencies. Not only are we not sure to what extent they are representative for the type of English that Dutch speakers are exposed to, but more importantly, they do not allow us to make a distinction between source-related and spontaneous use of the catchphrase for English. We will thus only be able to make definite conclusions on the interplay between both variables once we have collected more reliable information on the exposure to the expression in English, by searching more stable corpora.

Overall, we see how both familiarity with the original media source of a catchphrase and the overall exposure to the expression in (International) English are influential for the borrowability of catchphrases. We find a weak correlation between these features, which can only be reliably interpreted by conducting further analyses.

4.1.2 Sentence-type

For sentence-type, we find a clear effect on the borrowability of catchphrases. The declarative sentences are taken as reference point. No significant difference exists with questions and requests, but imperative and exclamative catchphrases do behave significantly different. Specifically, we find that these catchphrases have significantly less chance of occurring spontaneously in the corpus than declarative catchphrases.

This effect is tightly linked to the formal character of newspaper language: exclamations and commands are all but typical for this register. Hence, this type of catchphrases is less likely to occur in our corpus. What this factor shows us, then, is the importance of the characteristics of the receptor in issues of borrowability. Specifically, due to the formal nature of the “receiving” register, the use of exclamative and imperative catchphrases is hindered. To corroborate this claim, further analyses should be conducted, based on less formal language use.

4.1.3 Source-Type: Movies versus Series & Shows

The final and least important predictor in the model compares catchphrases from two SOURCE-TYPES, movies (as reference point) on the one hand and series and shows on the other. Table 4.1 indicates that the latter are less likely to occur spontaneously in our corpus than the former. Specifically, contingency tables for the feature show that 44% of the movie catchphrases are found spontaneously in the Dutch corpus, whereas this is true for only 16% of the catchphrases from series and shows.

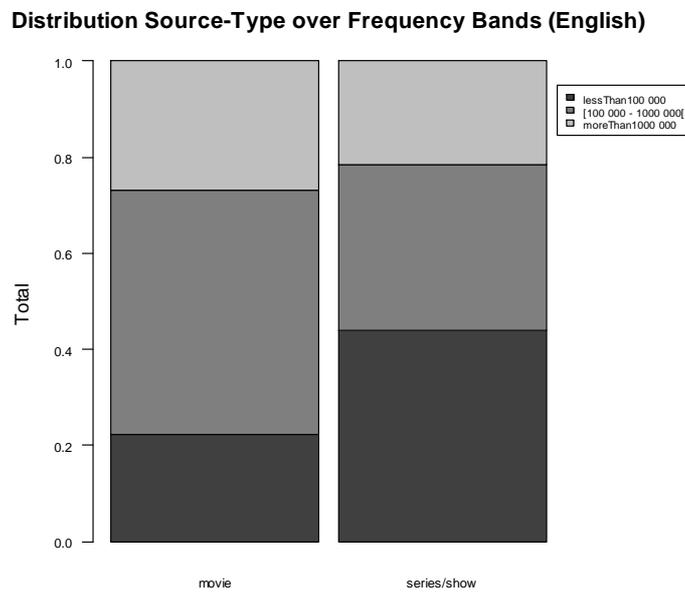


Figure 4.2

No clear-cut interpretation of this effect is available. A first important step in understanding the pattern is determining whether the effect of source-type is reliable or whether it should be seen as an artefact of the data collection. Specifically, we calculate the correlation of source-type with the entrenchment of the catchphrase in English, in order to verify whether it so happens that for movies, the online list makers

only incorporate popular catchphrases from entrenched sources, whereas for series, they select all types of catchphrases from a broader variety of sources.

A significant, but weak correlation is found between both predictors ($p = .01$, $\rho = -.17$). As is demonstrated by Figure 4.2, the effect is especially visible in the lowest frequency band (less than 100 000 Google hits for the catchphrase). Almost half of the catchphrases from series and shows are part of this category, whereas only about one fifth of the movie catchphrases is found here.

However, the correlation between both predictors does not completely explain the higher success-rate for movie-catchphrases. This is not only shown by the weak correlation coefficient, but also by the fact that both predictors occur side by side in the regression model, without this leading to multicollinearity. Furthermore, if the pattern found for source-type could be explained solely by the higher entrenchment of movie-catchphrases, the success-rate for movies and series should be similar when the frequency bands are held constant. However, Table 4.2 clearly shows that the higher success-rate for movies occurs consistently across the three frequency bands.

frequency-band (Google in English)	success-rate movies	success-rate series/shows
lessThan100 000	0.20	0.06
[100 000 - 1000 000[0.47	0.16
moreThan1000 000	0.61	0.23

Table 4.2

Apparently, the effect of source-type is not an artificial consequence of the nature of the database. The question then begs itself how we should interpret the difference between movies and series and what it can tell us about the nature of the English media that Dutch language users are exposed to. One possibility is to link the effect we see in Table 4.1 to a difference in cultural impact of movies and series. More specifically, movies can be said to have more cultural weight and thus a higher symbolic value, than series. This entails that the popularity of movie catchphrases might well be linked to the nature of the receptor corpus. A hypothesis is then that the success-rate for catchphrases from series and shows will be higher in colloquial language and informal registers (e.g. Usenet, IRC, youth language etc.). This claim will be verified in future research.

4.1.2 Insignificant predictors

Table 4.2 gives an overview of the predictors that did not make it to our final model. Two comments about this set have to be made.

Firstly, we would like to link the significance of SOURCE-RELATED USE to the absence of the alternative operationalization for media entrenchment (based on Google frequencies in Dutch). As both predictors are used as measures for the same feature, it is not remarkable that only one of them is selected in the model. Because the definition of source-related use is based on the same corpus as the response, this factor forms a more reliable and robust reflection of the influence of media entrenchment than the broader approach taken for the Google frequency counts. Hence, it is not surprising that this variable is selected for the model.

Predictor

source: media entrenchment (Google in Dutch)
source: country of origin (US/UK)
source: age
catchphrase: number of words
catchphrase: non-conventional vocab

Table 4.3

Secondly and more importantly, it is remarkable that we find no significant effect of the country of origin of the catchphrase. As we applied a less strict criterion for the initial selection of UK-catchphrases, we would expect this to show up in a lower success-rate for these observations. More specifically, because several of the UK-catchphrases were mentioned in only one online list, the catchphrase-status of these expressions is more dubious. Furthermore, as we have a limited set of UK-phrases (47 versus 182 US-phrases), these “dubious” catchphrases carry quite some weight in the analyses. Moreover, an often heard claim is that English influence is predominantly US-oriented. All of these observations lead us to expect a higher success-rate for US-catchphrases. Yet, this is not reflected in the data.

4.2 Lectal Variation: Belgian Dutch vs. Netherlandic Dutch

A final step in determining the influential factors for the borrowability of catchphrases in Dutch is finding out whether any lectal variation exists in the importance of the different factors. In this study, we focus on regional variation by determining whether any differences exist between the two main varieties of Dutch: Belgian Dutch (spoken in Flanders, the northern part of Belgium) and Netherlandic Dutch.

Previous research on the attitudes of both regions towards foreign influence leads us to believe that differences in the borrowability of catchphrases might exist. Specifically, Geeraerts *et al.* (1999) discuss the long-standing purist tradition in Flanders, which originated as a reaction against the French-speaking elite and then spread out to all forms of foreign influence, including English. However, Geeraerts *et al.* (1999) also note this purist tradition started declining since the second half of the twentieth century. The Netherlands, on the other hand, have always been characterised by an open attitude towards foreign languages in general, and towards English in particular.

In order to determine possible differences between the varieties concerning the use and borrowability of catchphrases, we compare the regression models of both regions. Firstly, we verify whether the same set of predictors is selected. Next, for the predictors occurring in both models, we compare the confidence intervals of the estimates in order to determine whether the effect-size is equal. The effect-size is significantly different when these intervals show no overlap. Before we discuss the results, we want to note that (1) no (striking) differences were found in the amount of catchphrases found in both corpora (2) both models showed a good fit.

The results of the comparison show that no differences between the regions can be noted. Firstly, as Table 4.2 illustrates, the same predictors are selected for both regions, in the same order. This means that the relative importance of the predictors is the same. Next, when comparing the confidence intervals, we find that no significant

differences exist in the effect-size of each of the predictors. This is reflected by the fact that the intervals for the regions show overlap for all predictors.

Coefficient	C.I. Belgian Dutch	C.I. Netherlandic Dutch	overlap?
SRU: yes	[2.07, 4.68]	[1.38, 3.26]	YES
exposure English: 1	[-0.23, 1.01]	[0.22, 1.45]	YES
exposure English: 2	[0.09, 0.80]	[-0.08, 0.57]	YES
sentence: question/request	[-0.47, 2.12]	[-1.76, 0.80]	YES
sentence: exclamation/command	[-2.10, 0.02]	[-2.09, -0.20]	YES
type: series/show	[-1.80, 0.08]	[-1.58, 0.12]	YES

Table 4.2

Hence, the results show that there are no striking differences in the use and borrowability of catchphrases between both regions. As such, this part of the analyses ties in with the results presented by Gerritsen *et al.* (2007). They also form further support for Geeraerts *et al.* (1999), who claim that, since the 1950s, the lexical differences between the regions have started to decline. Finally, it also ties in with the results discussed by Zenner *et al.* (2009b). This study, which inquires into the use of English job ads in Dutch and Flemish job ad magazines, shows that differences in the use of English between both regions have diminished since the 1970s and are now practically absent. The reason for this is a steep rise of the use of English in Flanders.

5. Conclusions & Prospects

The main point of this paper was to verify which factors influence the occurrence of spontaneously used English catchphrases in Dutch, with specific attention to the importance of media entrenchment. The regression analysis produced a stable model with quite some explanatory power, leading to four important points on the borrowability of catchphrases.

Firstly, we were able to indicate the importance of the original media origin of the catchphrase: direct media influence appears to play a role in the spread of catchphrases. Secondly, indirect influence was shown by the importance of the popularity of the catchphrase in International English. Thirdly, we indicated the importance of the characteristics of the investigated variety of the receptor language. Although no differences between the two main regional varieties (Belgian Dutch and Netherlandic Dutch) were visible, the results for sentence type and source type indicated the importance of register. Specifically, we saw how exclamative and commanding catchphrases are unsuccessful in our newspapers and how this is mainly due to the formal nature of the register. A similar argument was used to explain the higher success-rate for movie-catchphrases compared to catchphrases from series and shows. Finally, we found no difference between catchphrases from US and UK sources, although the US are presumed to be more important in the spread of English as an language for international communication.

To round off, we would like to present some prospects for future research. First of all, we want to get a better grasp on the distinction between direct media influence and indirect exposure to English. We will mainly develop this point by using a broader set of corpora for this research. We will firstly base the indirect exposure on more stable

corpora of English than Google (e.g. BNC, COCA and ELF-corpora). Secondly, we will incorporate Dutch corpora which represent lower registers.

Secondly, now that we have a better view on what determines *whether* we find catchphrases in our corpus, we can also focus on *how* we find them in our corpus. Three issues will be addressed. Firstly, we want to verify which types of catchphrases are subject to internal variation (compare Stefanowitch 2002). Secondly, we want to incorporate the distinction between catchphrases that are used spontaneously, but are still linked to their original source and catchphrases that are used completely freely (cf. *supra*). Finally, more research is needed on the specific textual functions of spontaneously used catchphrases (compare Androutsopoulos *in press*).

The final, and most important, prospect is the expansion of the issue of borrowed phraseology to other types of fixed expressions. Catchphrases formed an interesting starting point for the analysis, as they allowed us to investigate how important direct media-influence is in the usability of these catchphrases. Overcoming the issue of set-external proof by using methods parallel to Boers & Stengers (2008), we will be able to verify how the importance of media influence for catchphrases relates to its importance in the borrowability of non-media-related formulaic sequences. Only then will we be able to make definite claims on the importance of mass media in the spread of English.

Acknowledgements

Many thanks from Eline to Tom for his professional and cheerful help on the xml-parsing.

References

- Alexander, R.J. 1983. Catch phrases rule OK. Allusive puns analyzed. *Grazer Linguistische Studien* 20: 9-30.
- Androutsopoulos, J. *to appear*. English 'on top': Discourse functions of English resources in German mediascapes. *Sociolinguistic studies* (Norface special issue).
- Boers, F. & H. Stengers. 2007. A quantitative comparison of the English and Spanish repertoires of figurative idioms. In Boers, F. & S. Lindstromberg (eds.), *Cognitive Linguistic Approaches to Teaching Vocabulary and Phraseology*, 355-374. Berlin/New York: Mouton de Gruyter.
- Carstensen, B. 1965. *Englische Einflüsse auf die deutsche Sprache nach 1945*. Heidelberg: Carl Winter Universitätsverlag.
- Crystal, D. 2003. *English as a Global Language*. Second Edition. Cambridge: Cambridge University Press.
- De Sutter, G., D. Speelman & D. Geeraerts. 2008. Prosodic and syntactic-pragmatic mechanisms of grammatical variation: the impact of a postverbal constituent on the word order in Dutch clause final verb clusters. *International Journal of Corpus Linguistics* 13: 194-224.
- Deuchar, M., P. Muysken & S.-L. Wang. 2007. Structured variation in codeswitching: towards an empirically based typology of bilingual speech patterns. *The International Journal of Bilingual Education and Bilingualism* 10, 3, 298-339.

- Eckert, P. *under review*. Three waves of variation study: The emergence of meaning in the study of variation. <http://www.stanford.edu/~eckert/PDF/ThreeWavesofVariation.pdf>. 1 November 2009.
- Field, F.W. 2002. *Linguistic Borrowing in Bilingual Contexts*. Amsterdam: Benjamins.
- Filipovic, R. 1977. English words in European mouths and minds. *Folia Linguistica* XI, 3-4: 195-206.
- Fink, H. 1997. *Von Kuh-Look bis Fit for Fun: Anglizismen in der heutigen deutschen Allgemein- und Werbesprache*. Frankfurt am Main: Peter Lang.
- Geeraerts, D. 2005. Lectoral variation and empirical data in Cognitive Linguistics. In Ruiz de Mendoza Ibáñez, F.J. & M.S. Peña Cervel (eds.), *Cognitive Linguistics: Internal Dynamics and Interdisciplinary Interaction*, 163-189. Berlin/New York: Mouton de Gruyter.
- Geeraerts, D., S. Grondelaers & P. Bakema. 1994. *The Structure of Lexical Variation. Meaning, Naming and Context*. Berlin/New York: Mouton de Gruyter.
- Geeraerts, D., S. Grondelaers & D. Speelman. 1999. *Convergentie en divergentie in de Nederlandse woordenschat: een onderzoek naar kleding- en voetbaltermen*. Amsterdam: Meertens Instituut.
- Gerritsen, M., C. Nickerson, A. Van Hooft, F. Van Meurs, U. Nederstigt, M. Starren & R. Crijns. 2007. English in product advertisements in Belgium, France, Germany, the Netherlands and Spain. *World Englishes* 26, 3, 291-315.
- Grigg, P. 1997. Toubon or not Toubon: the influence of the English language in contemporary France. *English Studies* 4: 368-384.
- Haspelmath, M. 2008. Loanword typology: Steps toward a systematic cross-linguistic study of lexical borrowability. In Stolz, T., D. Bakker & R. Salas Palomo (eds.), *Aspects of Language Contact: New Theoretical, Methodological and Empirical Findings with Special Focus on Romancisation Processes*, 43-62. Berlin/New York: Mouton de Gruyter.
- Haugen, E. 1950. The analysis of linguistic borrowing. *Language* 26, 2: 210-231.
- Hilgendorf, S. 2007. English in Germany: Contact, spread and attitudes. *World Englishes* 26, 2, 131-148.
- Impe, L., Geeraerts, D. and Speelman, D. (2008). Mutual intelligibility of standard and regional Dutch language varieties. *International Journal of Humanities and Arts Computing* 2, 1-2, 101-117.
- Kachru, B.B. 1992. *The Other Tongue: English across Cultures*, second edition. Urbana: University of Illinois Press.
- Kowner, R. & S. Rosenhouse. 2008. The hegemony of English and determinants of borrowing from its vocabulary. In Kowner, R. & S. Rosenhouse (eds.), *Globally Speaking. Motives for Adopting English Vocabulary in Other Languages*, 4-18. Clevedon: Multilingual Matters.
- Krauss, P.G. 1958. The increasing use of English words in German. *The German Quarterly* 31, 4: 272-286.
- Kristiansen, G. & R. Dirven (eds.). 2008. *Cognitive Sociolinguistics: Language Variation, Cultural Models, Social Systems*. Berlin/New York: Mouton De Gruyter.
- Kurth, E-N. 1998. American idiom in Modern German: socio-linguistic motives. *Babel: Revue internationale de la traduction* 44, 3, 193-206.
- Labov, W. 2001. *Principles of Linguistic Change. Social Factors*. Oxford: Blackwell Publishers.
- Lathey, E. 1986. Pragmatic classification of idioms as an aid for the language learner. *IRAL, International Review of Applied Linguistics in Language Teaching* 24, 3: 217-233.
- Levshina, N., D. Geeraerts & D. Speelman. 2010. Towards a 3D-grammar: Variation of Dutch causative constructions. Pre-published paper for the *Laud Symposium on Cognitive Sociolinguistics*, Landau, Germany, 15-18 March 2010.

- Moon, R. 1998. *Fixed Expressions and Idioms in English*. Oxford: Clarendon Press.
- Muysken, P. 2000. *Bilingual Speech. A Typology of Code-Mixing*. Cambridge: CUP.
- Nettmann-Multanowska, K. 2003. *English Loanwords in Polish and German after 1945: Orthography and Morphology*. Frankfurt am Main: Peter Lang.
- Onysko, A. 2007. *Anglicisms in German. Borrowing, Lexical Productivity, and Written Codeswitching*. Berlin/New York: Walter de Gruyter.
- Pfaff, C. 1979. Constraints on language mixing: Intrasentential code-switching and Borrowing in Spanish/English. *Language* 55, 2: 291-318.
- Poplack, S. 1980. Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: toward a typology of code-switching. *Linguistics* 18, 581-618.
- Poplack, S., D. Sankoff & C. Miller. 1988. The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics* 26, 47-104.
- Ruette, T., D. Speelman & D. Geeraerts. 2010. "Tell me who you talk to, and I'll tell you how you talk": Comparing the language use of two interaction based clusters of people in a single Usenet newsgroup. Pre-published paper for the *Laud Symposium on Cognitive Sociolinguistics*, Landau, Germany, 15-18 March 2010.
- Seidlhofer, B. 2009. Common ground and different realities: World Englishes and English as a Lingua Franca. In *World Englishes* 28, 2, 236-245.
- Sharp, H. 2001. *English in Spoken Swedish. A Corpus Study of Two Discourse Domains*. Stockholm: Almqvist and Wiksell.
- Stefanowitsch, A. 2002. Nice to *miet* you: bilingual puns and the status of English in Germany. *Intercultural Communication Studies* XI, 4, 67-84.
- Yang, W. 1990. *Anglizismen im Deutschen*. Tübingen: Max Niemeyer Verlag.
- Yano, Y. 2009. English as an international lingua franca: from societal to individual. *World Englishes* 28, 2, 246-255.
- Van Hout, R. & P. Muysken. 1994. Modelling lexical borrowability. *Language Variation and Change* 6, 1, 39-62.
- Whitney, W.D. 1881. On mixture in language. *Transactions of the American Philosophical Association* 12: 1-26.
- Wray, A. & M.R. Perkins. 2000. The functions of formulaic language: an integrated model. In *Language & Communication* 20: 1-28.
- Zandvoort, R.W. 1964. *English in the Netherlands: a study in linguistic infiltration*. Groningen: Wolters.
- Zenner, E., D. Geeraerts & D. Speelman. 2009a. Expeditie Tussentaal: leeftijd, identiteit en context in "Expeditie Robinson". *Nederlandse Taalkunde* 14, 1, 26-44.
- Zenner, E., D. Geeraerts & D. Speelman. 2009b. Managing your corporate identity. An analysis of English job ads in the two main varieties of Dutch. Paper presented at the *SLE conference*, Lisbon, Portugal, 9-12 September 2009.