Across Languages and Cultures 16 (2), pp. 209–231 (2015)
DOI: 10.1556/Acr.16.2015.2.4

A CORPUS-BASED MULTIVARIATE ANALYSIS OF LINGUISTIC NORM-ADHERENCE IN AUDIOVISUAL AND WRITTEN TRANSLATION

LYNN PRIEELS,[1] ISABELLE DELAERE, KOEN PLEVOETS, GERT DE SUTTER

EQTIS, Empirical and Quantitative Translation and Interpreting Studies
Department of Translation, Interpreting and Communication, Ghent University
Groot-Brittanniëlaan 45, B-9000 GENT, Belgium
Phone: +32 9 33 11 930, Fax : +32 9 33 11 929,
E-mail: lynn.prieels@ugent.be
E-mail: isabelle.delaere@ugent.be
E-mail: koen.plevoets@ugent.be
E-mail: gert.desutter@ugent.be

Abstract: The present paper investigates linguistic norm-adherence in Belgian Dutch written and audiovisual translation. More particularly, it is measured to what extent language use in subtitles, in comparison to regular written translations and non-translations, conforms to explicit linguistic norms. Additionally, we analyze which effect different contextual parameters have on the extent of norm-adherence in Belgian Dutch subtitles. We use the Dutch Parallel Corpus and the SoNaR Corpus, and we analyze the data by means of profile-based correspondence analysis, yielding a visualization of norm-adherence distances between the different translation modes and non-translations. The results reveal that the parameters speaker type and source language significantly affect the degree of linguistic norm-adherence, whereas program genre has no influence. It is also shown that norm-adherence in subtitles holds a middle position between written translations and non-translations, which is explained in terms of target audience and communicative risk.

Keywords: audiovisual translation, norm-adherence, General Standard Dutch, Belgian Dutch

1. INTRODUCTION

The present paper is concerned with the question how Audiovisual Translation (AVT) deals with linguistic variation in the bicentric Dutch language area, which comprises both the Netherlands and the northern part of Belgium (Flanders).[2] Next to the official standard language that is shared by both Flanders and the Netherlands (General Standard Dutch), both areas are characterized by their

[1] The author holds a PhD fellowship with the Flemish Research Foundation (FWO).

[2] Apart from The Netherlands and Flanders, Dutch is also the official language in Suriname, Aruba, Curaçao and Saint-Martin.

own typical linguistic features that are widely used within that area, but that do not have the status of standard language. Hence, this study wants to investigate to what extent subtitlers in Flanders choose General Standard Dutch variants (accepted in both Flanders and the Netherlands) instead of non-general Belgian Dutch variants (which are widely used in Flanders, but do not have the status of general standard language) when subtitling English, Belgian Dutch and Netherlandic Dutch speech. This research question is particularly interesting, as our previous research has demonstrated that translators of 'regular' written genres generally opt more often for these General Standard Dutch words and constructions in comparison to writers of original texts (non-translators), which was assumed to be related to a standardizing, norm-adhering trend (Delaere et al. 2012; cf. also De Sutter et al. 2012 and Delaere and De Sutter 2013). Simultaneously, it was found that translators' behavior is not uniform at all, as norm-adherent lexical and grammatical choices were significantly dependent on extra-linguistic factors such as *source language*, *target audience* and *register* or *genre*: "registers with a lot of editorial control (fiction, non-fiction and journalistic texts) contain more standard language than the less edited registers (administrative texts and external communication)" (Delaere et al. 2012:203). This raises the question what kind of linguistic choices are made in Dutch subtitles in Flanders, considering the fact that they are (heavily edited) translations on the one hand (stimulating norm-adherent behavior) and written reproductions of spoken language, with its typical colloquial features, into a space-constrained written text on the other hand (possibly encouraging "colloquial", non-standard lexical choices) (Díaz-Cintas 2010:344–346; Karamitroglou 2000; Neves 2004).

The main goal of this article is to investigate (i) whether Belgian Dutch subtitles contain more or less General Standard Dutch linguistic features than non-general Belgian Dutch features in comparison with other translated and non-translated written genres, and (ii) whether the contextual parameters *program genre* (journalism, entertainment), *speaker type* (actor, voice-over) and *source language* (English, Netherlandic Dutch, Belgian Dutch) affect the linguistic choices in these subtitles. In order to obtain that goal, a varied list of 11 linguistic profiles (sets of synonymous linguistic items with at least one being General Standard Dutch and at least one being non-general Belgian Dutch), based on Delaere et al. (2012), is compiled. The profiles are subsequently extracted from the subtitle component of the SoNaR Corpus and the Dutch Parallel Corpus, and manually validated. The resulting frequency table is subjected to profile-based correspondence analysis (Plevoets 2008), which results in a visualization of the linguistic norm-adherence distances between the various components in our corpus, viz. Belgian Dutch subtitles and other written translations and non-translations.

This paper is structured as follows. Section 2 gives a rough sketch of the language situation in Belgium and its consequences for subtitling practice, which is necessary background information in order to understand the relevance of the selected profiles for the verification of the general norm-adhering idea central to this paper. As this paper is at the crossroads of Corpus-based Translation Studies and Audiovisual Translation Studies, Section 3 provides an overview of the current state of the art in both fields. The methodology underlying this empirical study is presented in Section 4, whereas Section 5 presents and discusses the obtained results. In the final section, we summarize the major conclusions ensuing from this investigation, and we point out directions for future research.

2. LANGUAGE SITUATION AND SUBTITLING PRACTICE IN FLANDERS

Together the Netherlands and the northern part of Belgium (henceforth: Flanders) constitute the main part of the Dutch speaking language area in the world, sharing the same standard language. However, because of specific historical developments, this standard language is not completely uniform in both parts of this language area, viz. there are some quite noticeable pronunciation differences in addition to lexical and grammatical differences. This situation is considered the result of a completely different standardization process. While the northern part of the Dutch language area started to develop a Dutch standard language (in the Renaissance), Flanders was politically, culturally and linguistically separated from the Netherlands. Instead of developing its own standard language, Flanders adopted French for supraregional communication. After the Dutch language was reinstated in Flanders, a gradual process that started in the late 19th century, and especially from the 1960s onwards, the language policy in Flanders was explicitly oriented towards the north: instead of further developing its own standard language, it was decided to adopt the standard language of the Netherlands in all its registers (Taeldeman 1992:42–44), making Netherlandic Dutch the dominant variety in the Dutch language area. As a consequence, language policy makers wanted to clear the language of the typical Belgian Dutch variants (that were considered dialectal or regiolectal) and have them replaced by the Netherlandic Dutch variants, which were assigned the status of General Standard Dutch, in order to create a common standard language. These attempts were partially successful in the more formal registers which have converged largely (but not completely) towards the northern variety (Geeraerts et al. 1999), but not in the informal registers. The remaining typical lexical and grammatical features of Belgian Dutch are nowadays tolerated to some extent (as the Belgian

Dutch variety is increasingly considered a variety next to the Netherlandic Dutch variety; De Caluwe 2011), but they are not fully accepted; consequently, given the historical dominance of the Netherlandic Dutch variety, professional writers still struggle with the status of the Belgian Dutch variants. In informal Belgian Dutch, the language policy was not successful at all, as there was a diverging tendency creating a considerable linguistic distance between the Netherlandic and Belgian Dutch colloquial variety on the one hand and between formal and informal Belgian Dutch on the other (cf. Geeraerts et al. 1999; Goossens 2000; Janssens and Marynissen 2008; Grondelaers and van Hout 2011).

These linguistic distances are reflected by the fact that Flemish and Netherlandic television programs are intralingually subtitled for viewers of the other part of the language area. Dialectal or regional speakers in Flanders, too, are increasingly subtitled in television programs intended for a general Flemish audience. Vandekerckhove et al. (2006, 2007) investigated this relatively new intralingual subtitling practice from an external-linguistic perspective. Using a corpus of 793 fiction programs, transmitted in 2005 by the Flemish public broadcasting company VRT and the commercial channel VTM, Vandekerckhove et al. were able to show that the Western regiolect was subtitled more often than the (dominant) Brabant regiolect on Flemish television. The Netherlandic Dutch variety is almost systematically subtitled. However, the question as to whether subtitles make more use of Belgian Dutch or General Standard Dutch still remains to be answered, which is the main goal of this paper.

3. OVERVIEW OF RELATED RESEARCH

Studies that focus on linguistic variability in audiovisual translation (AVT) are relatively scarce. Most of the attention in AVT went to the exploration of the general strategies that are used in AVT for coping with the information load in the original text. For dubbing, for example, Barambones Zubiria (2012) demonstrated that reduction and modulating are frequently used techniques to achieve isochrony between the source and target text utterances, whereas Remael (2007) shows that subtitles are frequently abbreviated in order to deal with the original speech tempo. On the other hand, Szarkowska (2008) found out that subtitlers were more explicit than the source text, for instance by inserting vocatives as a means to distinguish between different characters.

Nevertheless, there are a couple of recent studies in AVT that have started to explore specific linguistic characteristics in audiovisual translation. Baños (2013), for instance, found that the adverbial intensifiers *very*, *so*, *totally*, *pretty*, *really* in English-to-Spanish dubbed sitcoms are most frequently translated by

means of degree adverbs (e.g. *muy/tan* 'very' and *mucho/tanto* 'many'), making the dubbed speech less speech-like and more identical to written language. In her study on phrasal verbs in original Italian films and their French-to-Italian dubbed versions, Valentini (2013) argues that dubbed language is both lexically and grammatically poorer than original language, as the dubbed language contains fewer verb-particle constructions.

Other studies have tackled linguistic variability from a macro-perspective, viz. the rendering of geographically connotated language varieties. In the dubbing of English dialect into Italian, Ranzato (2010) distinguishes several strategies (e.g. rhyme, colloquial expressions and non-words) to create the effect of "an unlocalised variant of the standard language" (Ranzato 2010: 121). De Meo (2012) and Tortoriello (2012) both investigated the strategies used in translating Italian dialect into English subtitles. Those case studies demonstrate two different trends. Either the English subtitles were continuously written in standard English and did not make any attempt at conveying regional varieties (Tortoriello 2012), a technique which can be considered standardization, or the marked dialect intonation was translated by using non-standard grammatical constructions, loan words, idiomatic expressions, etc. (De Meo 2012).

The present study wants to contribute to this emerging trend within AVT studies to focus on variable linguistic choices, by adding a multivariate dimension (studying many linguistic features and many contextual parameters simultaneously) as well as by linking the research results more explicitly to well-known explanatory principles in Translation Studies, viz. standardization or normalization (for a discussion of the relationship between both principles, see, for example, Delaere et al. 2012). By doing so, we will be able to directly compare translation behavior in AVT with translation behavior in 'regular' written translations.

Previous research on written translations has indeed shown that, traditionally, there is a general tendency to normalize or standardize translations, which is for instance illustrated by the fact that more neutral expressions, more conventional and less creative language is used in translations, compared to their source texts or comparable non-translated texts. Nevertheless, recent research has repeatedly shown that this standardization tendency is just a tendency, and not a universal, as its occurrence significantly depends on contextual parameters as genre and source language. De Sutter et al. (2012) and Kruger and van Rooy (2012) have demonstrated this for the use of formal lexemes and structures in Dutch and South African English respectively, whereas Delaere et al. (2012) and Delaere and De Sutter (2013) came to the same conclusion on the basis of a case study on General Standard Dutch lexemes vs. (non-general) Belgian Standard Dutch lexemes.

4. METHODOLOGY

4.1. Hypotheses

Against the background of the information in Sections 2 and 3, we can formulate two hypotheses.

- Hypothesis 1: Flemish subtitlers are less norm-adhering than translators of regular written texts (i.e. subtitlers use less General Standard Dutch words and constructions vs. non-general Belgian Dutch items). There are three arguments underlying this hypothesis. First, their subtitles are mostly intended for a Flemish audience only, so the need to comply with the strictest linguistic norms of the dominant variety in the Dutch language area might be smaller in Flemish subtitles compared to regular written translations that are distributed in a larger area. Second, because the original auditory signal remains present, subtitlers are aware that their subtitles run a lower risk of being misunderstood (or not understood at all) compared to regular translations (where the original is replaced by the translations). Finally, the spoken source texts contain more traces of orality that triggers the use of non-general varieties.
- Hypothesis 2: the smaller norm-adhering tendency in subtitles (cf. Hypothesis 1) will be even more outspoken when spontaneous speech is subtitled (vs. voice-over speech) [Hypothesis 2a], in entertainment programs (vs. journalistic programs) [Hypothesis 2b] and in subtitling Belgian Dutch speakers (vs. speakers of Netherlandic Dutch and English) [Hypothesis 2c], as in these contexts subtitlers will be more exposed to Belgian Dutch variants, and hence are triggered more often to re-use these variants (which could be interpreted as a type of interference).

4.2. Corpus Materials

In order to support these hypotheses, we consulted two corpora. The first corpus is the SoNaR Corpus, a 500-million word balanced reference corpus for contemporary (1954–present) written Dutch (Reynaert et al. 2010), which is regionally (Belgian Dutch vs. Netherlandic Dutch) and stylistically stratified (36 text types, including newspapers, reports, emails, text messages and subtitles). For the present purposes, we obviously only selected the corpus component with subtitles created for Flemish television. More particularly, we opted to study the lexical norm-adherence behavior in 22 television programs that were subtitled for Flemish television between 2001 and 2005. We selected both jour-

nalistic and entertainment programs (genre difference), which were either interlingually (source language is English) or intralingually subtitled (source language is either Belgian Dutch or Netherlandic Dutch). The total number of words in this corpus component is somewhat more than two million (n = 2,048,480). In *Table 1*, an overview is given of the structure and the contents of the selected SoNaR Corpus components.

Table 1

Overview of the structure and word count of the selected components of the SoNaR Corpus

Journalistic subtitles	**Entertainment subtitles**
1,983,082	65,398

As can be seen, the distribution between journalistic and entertainment subtitles in the SoNaR Corpus is highly skewed: journalistic subtitles constitute almost 97% of the entire subtitles corpus component. Although this obviously is not an ideal distribution, it does not need to worry us too much, since the results of the statistical analyses to be performed in the remainder of this paper (cf. Section 4.4) are not affected by this skewed distribution.

Table 2

Overview of the structure and word count of the selected component of the Dutch Parallel Corpus

Genre	**Non-translated Dutch**	**Translated Dutch (< English)**	**Translated Dutch (< French)**
Administrative texts	428,391	237,579	339,826
Journalistic texts	483,714	295,039	272,429
Instructive texts	106,640	0	45,371
External communication	371,154	311,493	261,640
Literary texts	412,712	0	212,866
Total	1,802,611	844,111	1,132,132

The linguistic behavior in the subtitles component of the SoNaR Corpus is compared to the behavior of regular written translations and non-translations in the Dutch Parallel Corpus (DPC; Macken et al. 2011). DPC is a bidirectional parallel corpus with (Belgian and Netherlandic) Dutch as a source language (trans-

lated into French or English) and as a target language (translations from French or English). It consists of more than ten million words, it is stratified across five genres (literature, journalistic texts, administrative texts, instructive texts and texts for external communication), and it is sentence-aligned, part-of-speech tagged and lemmatized. For this study, we only selected translations and non-translations that were published in Flanders, and eliminated the texts of which the source language was unknown. *Table 2* gives an overview of the structure and size of the selected corpus components (total size: n = 3,778,854).

4.3. Data Extraction, Validation and Annotation

In order to trace linguistic norm-adherence in various translated and non-translated genres, we chart the frequency distributions of General Standard Dutch lexemes *in contrast to* their non-general counterparts. The basic idea underlying this approach is that norm-adherence can only be reliably measured if the proportion of General Dutch features is studied in combination with the proportion of their non-general Belgian Dutch linguistic alternatives. This is what Speelman et al. (2003) called the profile-based approach, each combination of a General Dutch variant and a Belgian Dutch variant being a profile (for advantages of this approach in the context of Translation Studies, see De Sutter et al. 2012:330–332).

A consequence of the profile-based approach is that only those General Dutch items can be studied which have a synonymous Belgian Dutch alternative. *Table 3* presents the 11 linguistic profiles that underlie the analyses in this study (cf. the appendix for a representative selection of corpus examples of each of the profiles). The list contains lexical profiles (#1, #6, #7, #9 and #11), grammatical profiles (#3, #8 and #10) and orthographic profiles (#2, #4, #5). The list of profiles is almost identical to the list in Delaere et al. (2012), where the norm-adhering hypothesis was studied for written translations vs. non-translations in isolation. The selection of profiles in Delaere et al. (2012) was based on a number of normative sources, such as the official website of the Dutch Language Union (http://taaladvies.net), which is currently considered the most important source of normative guidelines for language users in the entire Dutch language area, and Correct Taalgebruik (Correct language use) (Penninckx, Buyse and Smedts 2001). However, as the normative status of two Belgian Dutch lexemes has changed since Delaere et al. (2012), we chose to eliminate two profiles from the original list in Delaere et al. (2012) (*geraken-raken* ‘to get’; *bekomen-verkrijgen* ‘to obtain’).

Table 3
Overview of the linguistic profiles used in this study

Profile	General Standard Dutch	Non-general Belgian Dutch	Translation or meaning
1	*akkoord gaan met*	*akkoord zijn met*	*to agree with*
2	*een van de*	*één van de*	*one of the*
3	part + aux + inf aux + inf + part	aux + part + inf	order of the verbal end group
4	*te veel*	*teveel*	*too much*
5	*tenminste / ten minste* – correct context	*tenminste / ten minste* – wrong context	*at least / a minimum of*
6	*zulke* + plural noun	*zo'n* + plural noun	*such* + plural
7	*een beroep doen op*	*beroep doen op*	*to make an appeal to*
8	*proberen te* + inf	*proberen* + inf	*to try (to)* + inf
9	*op het eerste gezicht*	*op het eerste zicht*	*at first sight*
10	*beginnen te* + inf	*beginnen* + inf	*to start (to)* + inf
11	*zodra*	*van zodra*	*as soon as*

We extracted all variants in each of these profiles from both the SoNaR Corpus and the DPC by means of a custom search engine. All extracted data were then manually validated using a single, strict annotation guideline, which states that corpus attestations are only retained if they are synonymous with and can be replaced by their General Standard Dutch (GSD) or non-general Belgian Dutch (BD) alternative respectively. This resulted in a final dataset containing 7,067 relevant instances (SoNaR: n = 3,305; DPC: n = 3,762), the distribution of which across the profiles and the main language varieties is displayed in *Table 4*.

The relevant data from the DPC were already tagged for genre and source language. The data from the SoNaR Corpus had to be manually annotated for the contextual parameters *(program) genre* and *source language* on the one hand (in order to increase comparability with the DPC data) and *speaker type* on the other hand (in order to support more specific hypotheses on the effect of context on linguistic behavior in subtitles). The annotation of *genre* was partially based on the genre list in Van Gijsel et al. (2008:210–212), which makes a distinction between *journalistic television programs*, the main objective of which is to inform the audience (the typical example is a reportage, introduced by a presenter and/or commented on by a narrative voice) and *entertainment programs*, the main objective of which is to divert the audience (the typical

Table 4

Overview of the profile frequencies per language variety

Profile	Variants	Label	Non-translated written Dutch	Translated written Dutch < English	Translated written Dutch < French	Dutch subtitles
1	*akkoord gaan met*	GSD	45	11	30	32
	akkoord zijn met	BD	6	0	1	13
2	*een van de*	GSD	318	245	159	556
	één van de	BD	208	189	69	781
3	part + aux + inf aux + inf + part	GSD	29 33	16 19	19 59	91 41
	aux + part + inf	BD	27	3	3	20
4	*te veel*	GSD	123	36	74	216
	teveel	BD	25	4	5	14
5	*ten(minste)* correct	GSD	87	43	108	93
	ten(minste) wrong	BD	21	19	59	19
6	*zulke* + plural noun	GSD	57	39	19	173
	zo'n + plural noun	BD	3	0	0	11
7	*een beroep doen op*	GSD	87	29	108	31
	beroep doen op	BD	18	3	7	6
8	*proberen te* + inf	GSD	19	9	15	117
	proberen + inf	BD	2	1	2	1
9	*op het eerste gezicht*	GSD	13	10	4	32
	op het eerste zicht	BD	3	0	0	5
10	*beginnen te* + inf	GSD	14	6	10	54
	beginnen + inf	BD	9	2	4	26
11	*zodra*	GSD	93	37	103	225
	van zodra	BD	13	1	5	9

examples here are a docusoap or reality TV). The genre annotation of the television programs was done independently by the first and last author of this paper, and did not yield any conflicting annotations. The annotation of the source language was performed on the basis of the original footage, and turned out to be unproblematic: the source language was either Belgian Dutch, Netherlandic Dutch or English. The final contextual parameter is *speaker type*: on the basis of the speakers in the original footage, the subtitles were subdivided in *voice-over*

(either the presenter who introduces and concludes the program or the narrative voice who gives off-screen comments) or *actor/interviewee* (dialogic context).

4.4. Statistical Analysis: Profile-based Correspondence Analysis

We use profile-based correspondence analysis (Plevoets 2008) to analyze the associations between the profiles (rows) and the language varieties (columns; cf. *Table 4*) and to visualize these associations in a two-dimensional plot, as this has proved to be a particularly suitable method for this type of research (see for example Delaere et al. 2012; De Sutter et al. 2012; Delaere and De Sutter 2013). Profile-based correspondence analysis differs from standard correspondence analysis (Greenacre 2007) in that it is made sensitive of the profile structure in our data set. The two-dimensional plot will enable us to visually explore the lexical distances between the relevant translation modes (subtitles, written translations, written non-translations etc.). The basic rationale behind these kind of plots is that the distance between the language varieties (in this study for example, *translation modes*) is smaller if the proportions of the chosen linguistic variants for each of the 11 profiles in those varieties are similar. The position of these varieties relative to the position of the linguistic variants in the plot informs us of the linguistic options that are most often used in these varieties: the closer the varieties are to certain variants, the more often these variants are used in these varieties in comparison with the other varieties. For each of the language varieties, 95% confidence ellipses are calculated, which are the two-dimensional representations of the well-known 95% confidence intervals (Reiczigel 1996). If the confidence ellipses of two language varieties do not overlap, one can assert with 95% certainty that the distances between these varieties is statistically significant. In this respect, it is important to note that the size of the ellipses is also negatively correlated to the number of corpus instances underlying these ellipses (thus representing relatively larger statistical uncertainty). More detailed technical information on profile-based correspondence analysis and its advantages in the context of Translation Studies can be found in De Sutter et al. (2012) and Delaere et al. (2012). All analyses in this paper were done in R (R Development Core Team, 2014).

5. RESULTS AND DISCUSSION

In the present section, the results of the profile-based correspondence analysis are presented and discussed. In addition to providing some general information on the dispersion of the selected variants in the two-dimensional visualization,

which is to be considered the baseline for the interpretation of the linguistic behavior in the language varieties, Section 5.1 supports the first hypothesis, viz. that Flemish subtitlers are less norm-adhering than translators of regular written texts. In Sections 5.2 and 5.3, we support the second hypothesis, viz. that the contextual parameters *genre*, *source language* and *speaker type* have a significant effect on linguistc norm-adherence in subtitles.

5.1. Linguistic Norm-adherence in Different Translation Modes

In *Figure 1*, the distribution of the linguistic variants is visualized. The accepted General Standard Dutch (GSD) variants are marked in gray, the non-general Belgian Dutch (BD) variants are marked in black. If we look at the dispersion of these linguistic items in the plot, we can see that the accepted GSD variants are situated close to each other in the plot's origin, whereas the non-general BD variants are somewhat more widely distributed, but still mainly located in the fourth quadrant of the plot (bottom right), with both the items *proberen.inf* 'to try' and *akkoord.zijn* 'to agree with' as outliers. As correspondence analysis is basically a data reduction technique, which implies that some data is lost and that the representation in only two dimensions is merely an estimation of the associations in the original data frame, we also have to evaluate the quality of the visualization. As can be seen, the first dimension (x axis) accounts for 61.46% of the variation, whereas the second axis (y axis) captures 30.54%. This yields a total of 92%, so we can reliably state that the plot gives a very accurate visualization of the original variation.

In *Figure 2*, the different translation modes are plotted on to the individual variants, thus resulting in a bi-plot, which reveals two main findings. First, the linguistic choices made in each of the translation modes or language varieties are significantly different ($p < .05$), as the ellipses do not overlap. Second, and more specifically, if we look at the position of the different language varieties (ellipses) relative to the position of the linguistic variants, it can be seen that the variety that is most clearly related to non-general BD variants is original (non-translated) Dutch (text_DU_orig); it is the only variety that is located within the fourth (non-general) quadrant. In contrast, subtitles (AVT) are more clearly related to GSD variants, even though this variety is also surrounded by some of the non-general BD variants, whereas translations from English (text_DU<EN) are mostly related to accepted GSD variants, with the distance to the non-general variants being larger (compared to subtitles and non-translations). The most norm-adhering variety in our data, however, are translations from French (text_DU<FR), as the distance from this variety to the non-general BD variants

is the largest (and much larger compared to the distance to the accepted variants).

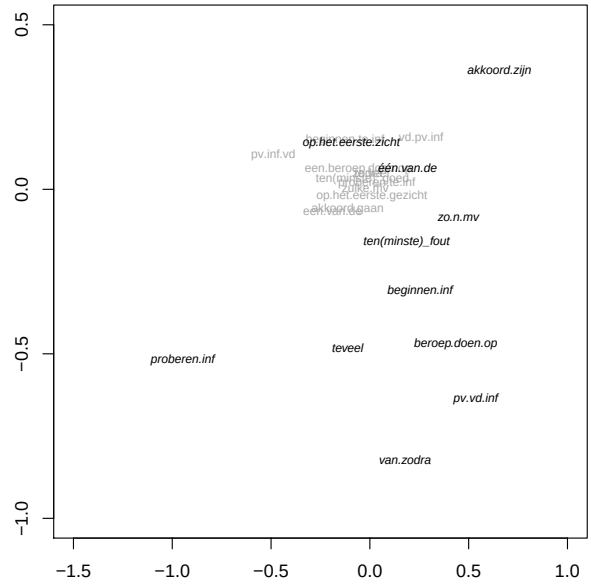

Figure 1. Distribution of the linguistic variants
(gray = GSD, black = non-general BD)

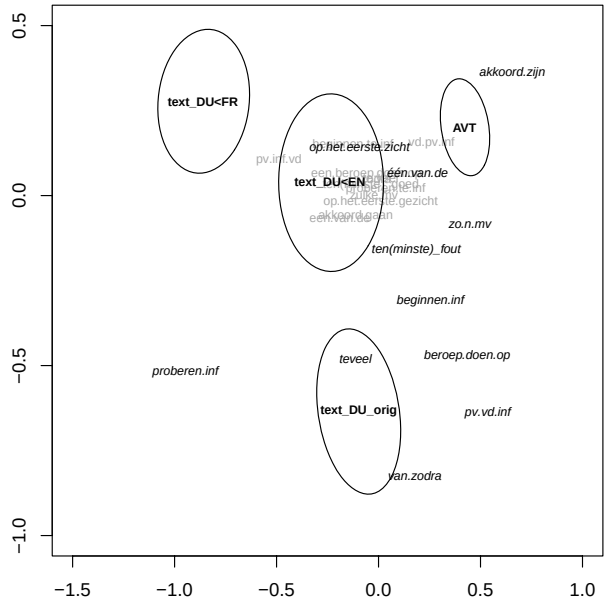

Figure 2. Bi-plot of the linguistic variants and translation modes
(gray = GSD, black = non-general BD)

The main conclusions to be drawn from these findings are the following. First, in line with our previous investigations (see Delaere et al. 2012:214–216), it can be re-confirmed that translations in general are more standardized or norm-adherent than non-translations, as they use GSD variants more frequently in comparison to non-translations. Second, translations do not behave uniformly, as source language (English, French) and translation mode (audiovisual, written) significantly affect the linguistic choices, and hence the degree of norm-adherence (the ellipses of the translated varieties do not overlap). Finally, and most importantly, the hypothesis that subtitles are less norm-adherent than regular written translations is also confirmed by the profile-based correspondence analysis, as subtitles are more related to non-general BD variants than the written translations (verification of Hypothesis 1). As mentioned in paragraph 4.1, we can see two main possible explanations for this. First, the subtitles in our data set are made by Flemish subtitlers working for Flemish television, and therefore the need to comply with the strictest linguistic norms (i.e. to use words and constructions from the dominant Netherlandic variety) is not as strong as for regular translators who mostly translate for a larger audience. Second, the communicative risk is relatively low, as the non-general BD variants occur very frequently in colloquial Belgian Dutch and the original speech remains available to the audience.

5.2. Contextual Parameters Influencing Linguistic Norm-adherence in Subtitles: Main Effects

Although the results in the previous section revealed that audiovisual translation (subtitles) in the Flemish context is less norm-adherent than regular written translations and more than non-translated texts, it is obvious that the analysis in 5.1 is rather coarse-grained, as it, for instance, does not take into account genre differences at all. In Delaere et al. (2012), we analyzed the effect of genre on norm-adherent behavior in Dutch translations and non-translations, which showed for example that journalistic texts, irrespective of them being translated or not, are more norm-adherent than instructive texts. For the Dutch subtitle data set, however, such an analysis has not yet been performed.

For that reason, this section analyzes the effect of the contextual parameters *program genre*, *source language* and *speaker type* on linguistic norm-adherence in Belgian Dutch subtitles. By doing so, we will be able to answer the question whether the language choices made in subtitles produced for Flemish television differ when an English speaker is subtitled (vs. a Belgian speaker vs. a Netherlandic speaker), when off-screen comments are subtitled (vs. on-screen speech) and when entertainment programs are subtitled (vs. journalistic programs) (cf. Hypotheses 2a, 2b, 2c).

As the effect of these parameters can only be measured for the subtitle part of our data set, we leave out the written translation data. Consequently, the positions of the linguistic variants and language varieties (ellipses) have to be recalculated for the subtitle data only. *Figures 3* and *4* present the outcome of this new profile-based correspondence analysis, which is again a very accurate two-dimensional representation of the original variation (91% = 70.20% + 21.13%).

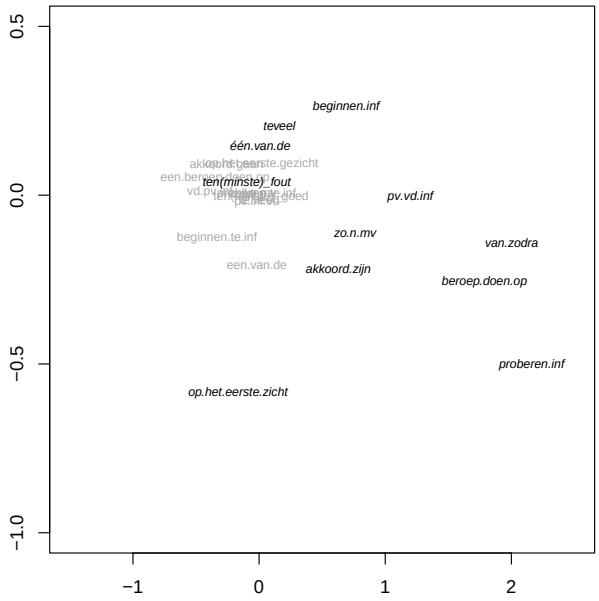

Figure 3. Distribution of the linguistic variants in the subtitle data set (gray = GSD, black = non-general BD)

Although the position of the variants indeed is somewhat different compared to *Figures 1* and *2*, the general picture more or less remains the same, with the accepted GSD variants located close to each other in the plot's origin, and the non-general BD variants more widely distributed mainly at the right-hand side of the plot. *Figure 4* shows the position of the different context-specific subtitles relative to the position of the linguistic variants. It becomes immediately clear that subtitles of actors' speech (actor) are most clearly related to non-general BD variants, whereas subtitles of off-screen voice-over comments (voiceover) are related to GSD variants. Subtitles that are intralingual translations of Belgian speakers' speech (the source language is then Belgian Dutch; intra.be) are also most clearly related to non-general BD variants, albeit less outspoken than subtitles of actors' speech (as intra.be is closer to the GSD variants than actor). These two types of subtitles are thus less norm-adhering than

the other types (verification of Hypotheses 2a and 2c), which are much more closer to the accepted, GSD variants: intralingual subtitles of Netherlandic Dutch speakers (intra.nl), interlingual subtitles of English speech (inter<EN), and subtitles in news and entertainment programs (news and entertainment). Within this group of norm-adhering subtitle contexts, we can see significant differences between subtitles in news programs, subtitles of voice-over speech and interlingual subtitles. There is also a significant difference between news programs and entertainment programs, but there are no significant differences between subtitles in entertainment programs and all other norm-adhering subtitle contexts on the one hand and between intralingual subtitles of Netherlandic speakers and all other norm-adhering subtitle contexts on the other hand.

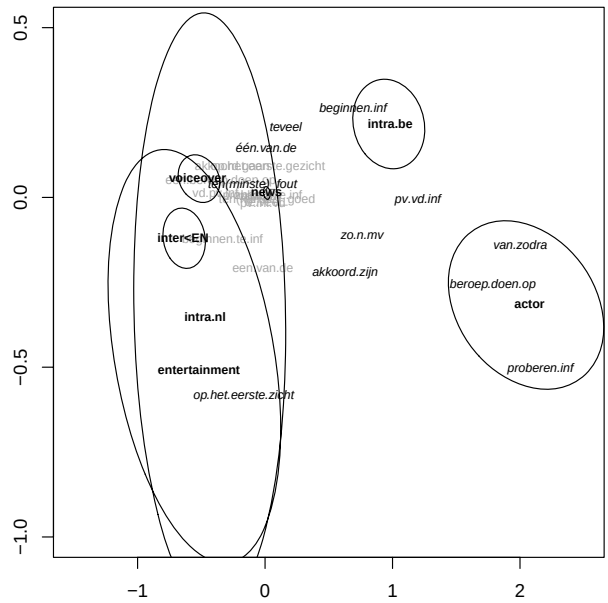

Figure 4. Bi-plot of the linguistic variants and the subtitle contexts (gray = GSD, black = non-general BD)

The lower degree of norm-adherence in intralingual subtitles of Belgian Dutch speakers as well as in subtitles of actors' speech (actor) can be explained by taking the nature of the original footage into account. First, actors' speech is, in contrast to voice-over speech (voiceover), mostly dialogic in nature (vs. monologic in voice-over speech), and thus has a greater chance of showing spontaneous, colloquial features. From that perspective, it seems plausible to suggest that this causes the increased frequency of non-general, but well-known and frequently attested colloquial BD variants in Flemish actors' speech. Sec-

ond, intralingual subtitles of Belgian Dutch speakers (intra.be) have a greater chance of containing BD variants than interlingual subtitles (inter<EN) or intralingual subtitles of Netherlandic Dutch speakers (intra.nl), as it is the only subtitle context in our data set where the subtitler is directly exposed to original Belgian Dutch speech. As a consequence, it seems safe to conclude that the lower level of norm-adherence in intralingual subtitles of Flemish speakers is reinforced (or even caused) by direct interference of the language use in the original Belgian Dutch television program (as we mentioned, the non-general BD variants occur very frequently in spoken language).

The significant differences between the subtitle contexts that are more closely related to the GSD variants are harder to interpret. It seems reasonable to state that subtitles in news programs (news) and subtitles of voice-over speech (voiceover) are equally norm-adherent, as they are similarly close to the GSD variants. The fact that their language behavior *does* differ significantly can be related to different linguistic preferences within the group of GSD variants. The ellipses of interlingual subtitles (inter<EN), intralingual subtitles of Netherlandic speech, and subtitles in entertainment programs are increasingly less related to the core of the GSD variants, as they all seem to move towards one of the BD variants, viz. *op het eerste zicht* 'at first sight'. This suggests that these subtitle contexts are norm-adhering to a large extent even though they are slightly more related to one of the non-general BD variants falsification of Hypothesis 2b.

In sum, the profile-based correspondence analysis of the subtitle data has confirmed that subtitles of actors' speech and Belgian Dutch speech are less norm-adhering than subtitles of voice-over speech, and interlingual subtitles and intralingual subtitles of Netherlandic speech respectively. On the other hand, it was only partly confirmed that subtitles in entertainment programs contain more BD variants than news programs. As mentioned, this is only true for one of the BD variants.

5.3. Contextual Parameters Influencing Linguistic Norm-adherence in Subtitles: Interaction Effects

The profile-based correspondence analysis in the previous section revealed interesting insights in norm-related linguistic behavior of certain specific subtitle contexts. What the analysis in Section 5.2 did not reveal, however, are the mutual relationships or interactions between these different subtitle contexts with respect to linguistic norm-adherence. What we do not know yet, for instance, is (i) whether Belgian actors are subtitled differently in comparison to Belgian voice-overs and (ii) whether voice-overs in entertainment programs are subtitled differently in comparison to voice-overs in news programs.

To answer those questions, we computed two-way interactions between *source language* and *speaker type*, between *source language* and *genre* and between *genre* and *speaker type*, and visualized these interactions in three bi-plots (*Figures 5*, *6* and *7*). The position of the linguistic variants in these interaction plots is unchanged compared to the main plot in *Figures 3* and *4*.

Figure 5 shows the interaction effect between the *speaker type* and *source language*. The most interesting observation is that intralingual subtitles of Belgian (Flemish) actors' speech (actor.intra.be) are clearly related to non-general BD variants, whereas intralingual subtitles of Belgian voice-over speech (voiceover intra.be) are located much closer to the GSD variants. In other words subtitles contain much more BD variants if the original speech is delivered by a Belgian Dutch speaker in a dialogic (colloquial) context compared to a monologic context. GSD variants are most frequently attested in interlingual subtitles of voice-over speech (voiceover.inter<EN) and, to a lesser extent, subtitles of Netherlandic actors' speech (actor.intra.nl). Interlingual subtitles of actors' speech (actor.inter<EN) is somewhat less norm-adherent, as it is closely related to one of the BD variants (*op het eerste zicht* 'at first sight'). All ellipses are significantly different from one another, except for actor.intra.nl, which overlaps with all other ellipses on the left-hand side of the plot.

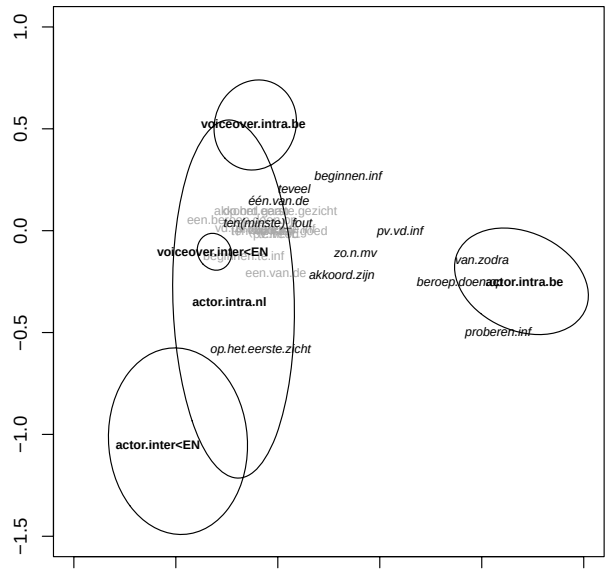

Figure 5. Bi-plot of the linguistic variants and the interaction between speaker type and source language in the subtitle data set (gray = GSD, black = non-general BD)

The main conclusion emerging from *Figure 5*, building on the relative distances between the ellipses, is that subtitles of Belgian Dutch actors' speech is the only subtitle type that is clearly characterized by non-general BD variants. We see two potential explanations for this, one in terms of interference or shining-through, one in terms of normalization. The first explanation would be that subtitlers transfer the BD variants from the spoken material into the subtitles, simply because this linguistic material is available in the original footage. This obviously can only be the case if it can be shown that the original speech of the Belgian Dutch actors contains more BD variants than all other types of speech (including Belgian Dutch voice-over speech). The second explanation would be that subtitlers strategically transfer these BD variants to create the spontaneous, colloquial style which is typical for dialogic contexts, and replace the BD variants by GSD variants in monologic contexts. This can only be the case if it can be shown that all Belgian Dutch source material (actors and voice overs) contains an equal amount of BD variants. As this study has not analyzed the original speech in detail, it is impossible to say which of the explanations is most plausible. Nevertheless, one could reasonably argue that the second explanation is less plausible (translators use BD variants strategically), given the situation in *Figure 5*, as we would then expect that all subtitles of actors' speech, irrespective of the source language, would contain much more BD variants than subtitles of voice-over speech – *quod non*. Obviously, follow-up research will have to be conducted in order to confirm that the first explanation is indeed most accurate (another explanation could be that subtitlers only use BD variants strategically if they are triggered by the original Belgian Dutch speech, but they do not insert them in contexts without a BD trigger).

Figure 6 shows the interaction between *source language* and *program genre*. There is only a clear significant difference between the linguistic choices in subtitles of Belgian speakers in news programs (intra.be.news) and all other subtitle types (intra.nl.news, inter<EN.entertainment, inter<EN.news). News programs in which Belgian Dutch speakers are subtitled are related most to the BD variants while interlingual and intralingual Netherlandic subtitles of news programs and interlingual subtitles of entertainment programs are located much closer to the GSD variants. This supports the interference explanation mentioned above, and shows that subtitlers do not strategically *add* BD variants in subtitles when the original footage does not contain BD variants at all.

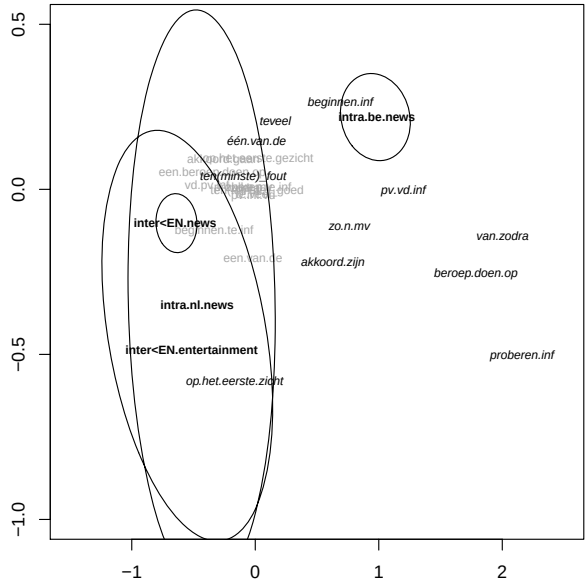

Figure 6. Bi-plot of the linguistic variants and the interaction between source language and program genre in the subtitle data set
(gray = GSD, black = non-general BD)

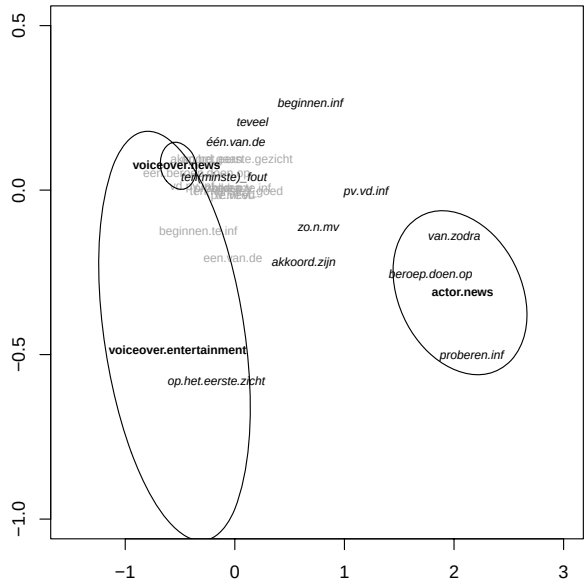

Figure 7. Bi-plot of the linguistic variants and the interaction between speaker type and program genre
(gray = General Standard Dutch, black = Belgian Dutch)

Finally, *Figure 7* presents the interaction between *speaker type* and *program genre*, in which it can be observed that subtitles of interviewees' speech in news programs (actor.news) contain much more BD variants than voice-overs' speech in news and entertainment programs, leading to the conclusion that BD variants most frequently show up in subtitles of spontaneous, dialogic speech.

6. CONCLUSIONS

Building on two large corpora of Belgian Dutch written and audiovisual translation, this paper analyzed how subtitlers, translators and original authors deal with norm-related language variation in the bicentric Dutch language area. More particularly, we performed several profile-based correspondence analyses of 11 linguistic profiles (or variables), consisting of at least one non-general Belgian Dutch variant and at least one accepted General Standard Dutch variant. The hypothesis that translators are more norm-adherent than non-translators was verified, but it was also shown that there is a considerable difference in linguistic use between audiovisual and written translation. Subtitle data contained significantly more non-general Belgian Dutch variants compared to regular written translations. In-depth analyses pointed out that linguistic choices in subtitles are mainly determined by the source language and by the speaker type. If the source language of the original footage is Belgian Dutch (yielding an intralingual translation), the amount of non-general Belgian variants increased significantly (compared to interlingual translations from English and intralingual translations from Netherlandic Dutch). Additionally, if the subtitled voice is of an actor or interviewee, the frequency of Belgian Dutch variants also increased significantly. The most obvious explanation offered for these results is that subtitlers (consciously or unconsciously) transfer the Belgian variants in the original footage directly to the subtitles, thereby maintaining the 'Belgian atmosphere' in the original footage. Further analyses will have to be performed to substantiate this explanation.

References

Baños, R. 2013. 'That is so Cool': Investigating the Translation of Adverbial Intensifiers in English–Spanish Dubbing through a Parallel Corpus of Sitcoms. In: *Perspectives: Studies in Translatology* Vol. 21. No. 4. 526–542.

Díaz-Cintas, J. 2003. Audiovisual Translation in the Third Millennium. In: Anderman, G. & Rogers, M. (eds) *Translation Today: Trends in Perspectives*. Clevedon: Multilingual Matters. 192–204.

Díaz-Cintas 2010. Subtitling. In: Gambier, Y. & van Doorslaer, L. (eds) *Handbook of Translation Studies 1*. Amsterdam & Philadelphia: Benjamins. 344–349.

Delaere, I., G. De Sutter & K. Plevoets 2012. Is Translated Language more Standardized than Non-translated language? Using Profile-based Correspondence Analysis for Measuring Linguistic Distances between Language Varieties. *Target* Vol. 24. No. 2. 203–224.

Delaere, I., & De Sutter, G. 2013. Applying a Multidimensional, Register-sensitive Approach to Visualize Normalization in Translated and Non-translated Dutch. In: Lefer M. A. & Vogeleer, S. (eds) Interference and Normalisation in Genre-controlled Multilingual Corpora. *Belgian Journal of Linguistics.* Vol. 27. 43–60.

De Meo, M. 2012. Subtitling Dialects: Strategies of Socio-cultural Transfer from Italian into English. In: Bruti, S. & Di Giovanni, E. (eds) *Audiovisual Translation across Europe. An Ever-changing Landscape. New Trends in Translation Studies* Vol. 7. Oxford e.a.: Peter Lang. 79–96.

De Sutter, G., Delaere, I., & Plevoets, K. 2012. Lexical Lectometry in Corpus-based Translation Studies. Combining Profile-based Correspondence Analysis and Logistic Regression Modeling. In: Oakes, M. & Ji, M. (eds) *Quantitative Methods in Corpus-based Translation Studies. A Practical Guide to Descriptive Translation Research.* Amsterdam/Philadelphia: John Benjamins.

Hendrickx, R. 1998. Het taalcharter. VRTtaal.net. Available on <http://www.vrt.be/taal/taalcharter>

Karamitroglou, F. 2000. *Towards a Methodology for the Investigation of Norms in Audiovisual Translation*. Amsterdam: Rodopi.

Kruger, H. & van Rooy, B. 2012. Register and the Features of Translated Language. *Across Languages and Cultures* Vol. 13. No 1. 33–65.

Macken, L., De Clercq, O. & Paulussen, H. 2011. Dutch Parallel Corpus: A Balanced, Copyright-cleared Parallel Corpus. *Meta* Vol. 56. No. 2. 374–390.

Neves, J. 2004. Language Awareness through Training in Subtitling. In: Orero, P. (ed.) *Topics in Audiovisual Translation*. Amsterdam & Philadelphia: John Benjamins. 127–140.

Penninckx, W., Buyse, P. & Smedts, W. 2001. *Correct Taalgebruik*. Kortrijk-Heule: UGA.

Plevoets, K. 2008. *Tussen spreek-en standaardtaal: een corpusgebaseerd onderzoek naar de situationele, regionale en sociale verspreiding van enkele morfo-syntactische verschijnselen uit het gesproken Belgisch-Nederlands.* Unpublished dissertation, Katholieke Universiteit Leuven.

Pym, A. 2008. On Toury's Law of how Translators Translate. In: Pym, A., Schlesinger, M & Simeoni, D. (eds) *Beyond Descriptive Translation Studies. Investigations in Homage to Gideon Toury*. Amsterdam & Philadelphia: John Benjamins. 311–328.

R Development Core Team 2014. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://R-project.org.

Ranzato, I. 2010. Localising Cockney: Translating Dialect into Italian. In: Díaz-Cintas, J. Matamala, A. & Neves, J. (eds) *New Insights into Audiovisual Translation and Media Accessibility.* Amsterdam/New York: Rodopi.

Remael, A. 2007. Vertaling in beeld: aspecten van audiovisuele vertaling en ondertiteling. In: Pauwels, L. (ed.) *Methodisch kijken: aspecten van onderzoek naar film- en beeldcultuur.* Leuven: Acco. 183–205.

Remael, A. & De Houwer., A. & Vandekerckhove, R. 2008. Intralingual Open Subtitling in Flanders: Audiovisual Translation, Linguistic Variation and Audience Needs. *Journal of Specialized Translation* No. 10. 76–105.

Reynaert, M., Oostdijk, N., De Clercq, O., van den Heuvel, H., & F. de Jong 2010. Balancing SoNaR: IPR versus processing issues in a 500-million-word written Dutch Reference Corpus. In: *Seventh conference on International Language Resources and Evaluation*, LREC '10, 19–21 May 2010, Malta. 2693–2698.

Taeldeman, J. 1992. Welk Nederlands voor Vlamingen? *Nederlands van nu* No. 40. 33–52.

Tortoriello, A. 2012. Lost in Subtitling? The Case of Geographically Connotated Language. In: Bruti, S. & Di Giovanni, E. (eds) *Audiovisual Translation across Europe. An Ever-changing Landscape. New Trends in Translation Studies* Vol. 7. Oxford e.a.: Peter Lang. 97–112.

Valentini, C. 2013. Phrasal Verbs in Italian Dubbed Dialogues: a Multimedia Corpus-based Study. *Perspectives: Studies in Translatology* Vol. 21. No. 4. 543–562.

Vandekerckhove, R., A. De Houwer, Remael, A. & Van der Niepen, I. 2006. Intralinguale ondertiteling van Nederlandstalige televisieprogramma's in Vlaanderen: linguïstische en extralinguïstische determinanten. In: Koole, T. Nortier, J. & Tahitu. B. (eds) *Artikelen van de vijfde sociolinguïstische conferentie (Anéla – Nederlandse vereniging voor toegepaste taalwetenschap)*. Delft: Eburon. 503–513.

Vandekerckhove, R. & De Houwer, A. & Remael, A. 2007. Intralinguale ondertiteling op de Vlaamse televisie: een spiegel voor de taalverhoudingen in Vlaanderen? Sandra, D., Rymenans. P., Van Petegen, P., (eds) *Tussen taal, spelling en onderwijs. Essays bij het emiraat van Frans Daems.* Gent: Academia Press.

Van Gijsel, S., Speelman, D., & Geeraerts, D. 2008. Style Shifting in Commercials. *Journal of Pragmatics* Vol. 40. No. 2. 205–226.