

# Between language policy and language reality: a corpus-based multivariate study of the interlingual and intralingual subtitling practice in Flanders

Lynn Prieels & Gert De Sutter

To cite this article: Lynn Prieels & Gert De Sutter (2017): Between language policy and language reality: a corpus-based multivariate study of the interlingual and intralingual subtitling practice in Flanders, Perspectives, DOI: [10.1080/0907676X.2017.1380053](https://doi.org/10.1080/0907676X.2017.1380053)

To link to this article: <http://dx.doi.org/10.1080/0907676X.2017.1380053>

 View supplementary material [↗](#)

 Published online: 10 Oct 2017.

 Submit your article to this journal [↗](#)

 View related articles [↗](#)

 View Crossmark data [↗](#)



# Between language policy and language reality: a corpus-based multivariate study of the interlingual and intralingual subtitling practice in Flanders

Lynn Prieels<sup>†</sup> and Gert De Sutter

Research Unit EQTIS (Empirical and Quantitative Translation and Interpreting Studies), Department of Translation, Interpreting and Communication, Ghent University, Ghent, Belgium

## ABSTRACT

The present study explores how subtitlers in Dutch-speaking Belgium deal with the linguistic tension between the norm-adherent language policy of the broadcaster they work for (VRT) – which is oriented towards the use of Belgian Standard Dutch – and the particular language situation in Flanders, which is characterized by the increasing use of Colloquial Belgian Dutch (CBD) in spoken contexts. Subtitlers, who produce written reproductions of spoken language, therefore need to mediate between the official language policy and language reality. In this context, this study aims to measure the extent to which the language used in Flemish subtitling conforms to the official language policy. Additionally, we analyze which contextual parameters affect the subtitlers' linguistic choices. The data are extracted from the SoNaR corpus and subjected to profile-based correspondence analysis, visualizing the linguistic behavior, and hence the degree of norm conformity, in the subtitles. The results reveal that subtitles on Flemish television are norm-adherent to a large extent, although certain contextual parameters (source language and program genre) enhance the use of (nonstandard) CBD. These results evidence the spread of CBD features from exclusively spoken registers to a written register that was, until recently, under heavy normative control.

## ARTICLE HISTORY

Received 31 May 2016  
Accepted 6 September 2017

## KEYWORDS

Audiovisual translation; subtitling; corpus-based research; norm-adherence; translation studies

## 1. Introduction

In the past few decades, the study of language variation in subtitling has been a prevalent topic in various linguistic-oriented studies, yielding many interesting insights. Cavalheiro (2008), for instance, analyzed the television subtitling of the film *Gone with the Wind*, showing that the spoken substandard variety was translated into an 'equivalent' Portuguese variety in subtitles on the private television channel. Other studies have demonstrated that nonstandard language varieties (such as dialect, slang, regiolect) in the spoken source text are generally standardized (i.e. interlingually or intralingually translated into standard language) in the corresponding subtitles. Not only are these

**CONTACT** Lynn Prieels  Lynn.Prieels@UGent.be

<sup>†</sup>The first author holds a PhD fellowship with the Flemish Research Foundation (FWO).

 Supplemental data for this article can be accessed at <https://doi.org/10.1080/0907676X.2017.1380053>

© 2017 Informa UK Limited, trading as Taylor & Francis Group

nonstandard linguistic features difficult to reproduce in written language, subtitlers are also frequently bound by the language policy of the subtitle authorities, which generally support a norm-adhering, standardizing tendency (e.g. Hamaida, 2007; Rемаel, De Houwer, & Vandekerckhove, 2008; Rosa, 2001). Although these studies offer a valuable first insight into language variation in subtitling, they suffer from a small empirical base (the subtitle material was collected from, respectively, two films, one film and four episodes selected from four different series). Moreover, they do not take into consideration the influence of contextual factors, such as register, on the language used in subtitling, which is crucial, as previous research has already demonstrated the relevance of these factors in language variation studies (e.g. Delaere, De Sutter, & Plevoets, 2012). Furthermore, the aforementioned studies merely apply a qualitative approach, without involving statistical techniques to analyze linguistic variation in the subtitles, which makes it impossible to reliably discover how different language varieties are related to each other and which factors determine subtitlers' linguistic choices. This study, on the contrary, quantitatively investigates linguistic variation in subtitling on Flemish television, using a large corpus of interlingual and intralingual Belgian-Dutch subtitles and examining subtitlers' linguistic choices in various contexts. The specific language situation in the Dutch-speaking part of Belgium, also called Flanders, makes this research particularly interesting. In Flanders, Belgian Standard Dutch (*Belgisch Standaardnederlands*, hereinafter BSD) is the official variety that is generally accepted and especially used in very formal spoken registers and in written language. In less formal and informal contexts, however, language use in Flanders is strongly characterized by typical lexical and grammatical features that are widely used, but not accepted as BSD by the language authorities (e.g. Dutch Language Union<sup>1</sup>, Hendrickx, 1998). This (nonstandard) informal variety is known as *tussentaal* (literally: 'in-between language'), or, as we will name it, Colloquial Belgian Dutch (CBD). The particular Flemish linguistic situation has driven Dutch-speaking public media to develop a language policy, with specific guidelines for television and radio hosts. The public broadcaster VRT<sup>2</sup>, for instance, adopts a clear norm-adherent policy, aiming to be 'the norm for the Belgian variety of the Dutch standard language' and 'therefore adopting an attractive, clear and correct standard language that takes into account and is adjusted to its audience' (authors' translation) (Hendrickx, 1998, p. 1). According to these guidelines, the public broadcaster occasionally allows the use of spoken nonstandard, regional-tinted varieties, such as CBD, in order to preserve the authenticity of the program; however, the default language variety to be used, especially in informative programs, such as news and documentaries, is BSD. Nevertheless, previous research has shown that nonstandard varieties (e.g. CBD, regiolect, dialect) are frequently spoken on Flemish television, even in programs that would actually require BSD (Van Hoof, 2010). This raises the question as to whether these nonstandard language varieties also appear in the subtitles, given 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 (possibly encouraging the use of nonstandard linguistic items) on the other (Diaz-Cintas, 2010, pp. 344–346; Karamitroglou, 2000; Neves, 2004). Thus, the main goal of this study is to investigate how Flemish subtitlers deal with this linguistic tension between the norm-adherent language policy on the one hand and the language reality on the other. In particular, this study examines (i) to what extent Flemish subtitlers use nonstandard colloquial linguistic items, (ii) whether the subtitles contain more colloquial lexemes

than colloquial grammatical constructions, (iii) which contextual parameters (*purpose, target audience, cast, ...*) affect these linguistic choices, and (iv) how these parameters are related to each other. In order to obtain these research goals, lists of lexical and grammatical profiles were compiled. These are sets of synonymous linguistic items with at least one being a BSD variant and at least one being a CBD variant. All these profiles were then extracted from the subtitle component of the SoNaR corpus and manually validated. In a last step, the resulting frequency table was subjected to profile-based correspondence analysis (Plevoets, 2008, 2015; for more technical information about this technique and its advantages we refer to De Sutter, Delaere, & Plevoets, 2012). This multivariate, statistical technique allows us to analyze the linguistic discrepancies in the language behavior of the subtitlers related to different linguistic contexts, and visualize them in a two-dimensional plot, which makes interpretation of the resulting patterns easier.

This paper is structured as follows. Section 2 first describes the language situation in the Dutch language area and its consequences for the subtitling policy in Flanders. Section 3 presents the data and the methodology of this corpus study; section 4 discusses the obtained results. Finally, in section 5, we summarize the major conclusions and elaborate on directions for future research.

## 2. Language situation and subtitling policy in Flanders

Due to particular historical developments (cf. Janssens & Marynissen, 2008; Willemyns & Daniëls, 2003), the Dutch language area is nowadays characterized by a complex language situation, in which both Flanders (in Belgium) and the Netherlands share the same supraregional standard language as well as their own area-specific standard language, viz. Netherlandic Standard Dutch and BSD, which are largely similar. This BSD variety is also called VRT Dutch by its users, referring to the language variety that is used in informative radio and television programs on the Flemish public broadcaster (Geeraerts, 1998, 2001). In informal registers, a considerable linguistic gap can be observed between Colloquial Netherlandic and Belgian Dutch on the one hand, and BSD and CBD on the other (cf. Geeraerts, Grondelaers, & Speelman, 1999; Goossens, 2000; Grondelaers & Van Hout, 2011; Janssens & Marynissen, 2008). CBD is characterized by lexical, grammatical and phonological features that are widely used in Flanders, but they are not accepted as belonging to BSD by the language authorities (e.g. Dutch Language Union, lexicographers). Nevertheless, several studies have demonstrated that this CBD variety has become increasingly prevalent on Flemish television (e.g. Lefevere, 2011; Prieels, 2013; Van Hoof & Vande kerckhove, 2013), as a consequence of which speakers of CBD (and other nonstandard varieties) are increasingly being subtitled on Flemish television (cf. Remael et al., 2008; Vandekerckhove, De Houwer, Remael, & Van der Niepen, 2006, 2007, 2009).

This linguistic tension is not only reflected in the increased use of intralingual subtitles on Flemish television, but also forces professional writers (including translators and subtitlers) to continuously evaluate the status of words, constructions and idioms that include those frequently used, but not accepted, CBD features. In a recent study, Delaere et al. (2013) revealed that translators generally tend to use more BSD words and constructions compared to writers of original texts (Delaere et al., 2013; cf. also De Sutter et al., 2012 and Delaere & De Sutter, 2013). Moreover, these linguistic choices have been demonstrated to

be significantly dependent on contextual factors such as source language, target audience and register.

The subtitling policy on Flemish television is largely oriented toward the general well-being of its audience. In an inquiry set up by the public broadcaster VRT in 2000, the main users of closed subtitling (the deaf and hard-of-hearing) declared that nonstandard utterances should be rendered in the subtitles to retain the authenticity of the program (Doens, 2000; Slembrouck & Van Herreweghe, 2004). However, this requirement contravenes the prevailing norm of the public broadcaster, which is oriented toward the general use of BSD on television (Hendrickx, 1998). The results of the study were published (Doens, 2000) and used for the compilation of the *Stijlboek voor Teletekstondertiteling* [*Style Guide for Teletext Subtitling*], a book with guidelines for closed subtitling on the public broadcaster. Next to guidelines on layout and the position of subtitles, this book contains one paragraph consisting of a small number of guidelines concerning the use of standard language and nonstandard, colloquial language. According to that style guide, subtitlers are expected to use BSD in the subtitles, thereby going against the audience's needs. Nevertheless, VRT occasionally leaves room for colloquial lexical items, which 'can be "more or less" reproduced in the subtitles, except for the colloquial forms of the personal pronoun (*ge/gij/uw* [you(r)]) and the flexion of articles (e.g. *nen* [a(n)]) and pronouns (e.g. *mijnen* [my])' (authors' translation) (Dewulf & Saerens, 2000, p. 35)<sup>3</sup>. Furthermore, editorial consultations are necessary to decide whether the nonstandard lexical item is acceptable in the subtitles. The guidebook does not contain, however, any guidelines concerning the use of nonstandard grammatical items. In other words, the Flemish public broadcaster principally adopts an attitude of what Rosa (1994, 2001) names 'centralization', i.e. translating nonstandard language into standard language, while ignoring the features of spoken verbal language. According to Rosa (2001), this strategy of centralization is mostly adopted by the public broadcaster in an attempt to uphold the standard variety and to transfer the prestige of the written norm (see also Cavalheiro, 2008; Pinto, 2009). Prieels, Delaere, Plevoets, and De Sutter (2015) indeed demonstrated that subtitles on Flemish television contain standard language to a large extent. Nevertheless, they also provided an initial indication that subtitles do not exclusively contain standard variants. In this study, building on the Dutch Parallel Corpus and the SoNaR corpus, the effect of different contexts on the subtitling practice on Flemish television was examined. It was demonstrated that Belgian-Dutch subtitles hold a middle position between translated and original Belgian-Dutch texts, as subtitlers use less BSD than translators do, but subtitles contain more BSD compared to nontranslated texts<sup>4</sup>. The study also revealed that subtitlers' linguistic choices are largely influenced by two contextual factors: the source language and the speaker's type. On the one hand, the number of CBD words and constructions significantly increased in intralingual subtitles of Belgian-Dutch speakers compared to interlingual subtitles of English speakers and intralingual subtitles of Netherlandic-Dutch speakers (cf. also Remael et al. (2008) and De Ridder (2015)). On the other, if the subtitled speech was produced by an actor or interviewee (instead of a voice-over), the frequency of CBD features also increased significantly. It can thus be assumed that Flemish subtitlers (consciously or unconsciously) transfer the nonstandard variants in the original footage more directly to the subtitles (Prieels et al., 2015, p. 229). These findings raise new questions about subtitling practices in Flanders. As previous research has focused predominantly on lexical features of standard and nonstandard

language, and VRT's subtitle guidelines only mention some tolerance towards the use of colloquial lexicon (whereas the use of colloquial grammatical constructions is left implicit), the question arises as to whether this means that subtitles indeed contain more colloquial lexemes than colloquial grammatical constructions. In other words, what are the exact proportions of lexical vs. grammatical standard and nonstandard features? Furthermore, the influence of other contextual parameters (*program genre, purpose, target audience, cast*) needs to be examined.

### 3. Methodology

First, we compiled three sets of linguistic profiles: one lexical set and two types of grammatical sets; and extracted them from our corpus (3.2.). In a next step (3.3.), the extracted data were manually validated and annotated for five language-external parameters (*source language, program genre, purpose, target audience* and *cast*). Finally, we applied a multivariate statistical technique called *profile-based correspondence analysis* (Plevoets, 2015) to measure and visualize the relationship between these external parameters on the one hand, and the use of BSD and CBD on the other (3.4.). In the next paragraph (3.1.), we will first set up the hypotheses that reflect our initial expectations.

#### 3.1. Hypotheses

Against the background of the subtitling practice on Flemish television described in Section 2, we can formulate the following hypotheses:

Hypothesis 1: Subtitles on Flemish television contain fewer CBD lexemes than CBD grammatical constructions.

Although its subtitle guidelines prescribe that the public broadcaster allows the reproduction of colloquial lexemes in the subtitles to some extent, it can be expected that subtitles on Flemish television contain more CBD grammatical constructions than CBD lexemes. Not only are subtitlers more aware of standard and colloquial lexemes, since lexical features are more salient than grammatical features (Lybaert, 2014), but the subtitle guidelines of the Flemish public broadcaster give the impression that the sensitivity toward the use of standard and colloquial lexemes is remarkably higher (Dewulf & Saerens, 2000). As a consequence, it can be expected that colloquial lexical variants are restricted to a minimum by the subtitlers.

Hypothesis 2: The use of standard and colloquial grammatical constructions is less context dependent compared to the use of standard and colloquial lexemes.

Besides the assumption that grammatical colloquialisms occur more frequently in Flemish subtitles than lexical colloquialisms, it can also be expected that these CBD grammatical items are not tied to specific contexts. In a study by Lybaert (2014), it was shown that syntactical features are less salient than lexical features are, which can be attributed to the abstract nature of the syntactical elements (Van Bree, 2000): unlike the lexicon, this domain is characterized by abstract rules, which makes it more automated or less concrete, so that language users unconsciously use and perceive these grammatical constructions. As a consequence, these grammatical features frequently occur in various situations in daily language use without being perceived very consciously; therefore, it can be expected that these elements are also frequently reproduced in different subtitle contexts.

Hypothesis 3: The number of CBD variants in the subtitles will be higher (a) in entertainment, infotainment and humor programs (vs. informative programs), (b) in programs with a general audience (vs. children), (c) when the spontaneous speech of nonactors (vs. scripted language of actors) is subtitled, and (d) when Belgian-Dutch speakers (vs. speakers of Netherlandic Dutch and English) are subtitled.

Not only does the public broadcaster VRT allow more language variation in these programs (Hendrickx, 1998), but subtitlers are more likely to be exposed to colloquial variants, which increases the odds that they reuse these variants in the subtitles (Prieels et al., 2015).

### 3.2. Corpus data

The data used for this study were extracted from the SoNaR corpus, a 500-million word balanced reference corpus for contemporary (1954–present) written Dutch (Reynaert, Oostdijk, De Clercq, van den Heuvel, & de Jong, 2010), which is regionally (Belgian Dutch vs. Netherlandic Dutch) and stylistically (36 text types, including newspapers, reports, emails, text messages and subtitles) stratified. Obviously, to analyze the linguistic behavior in Belgian-Dutch subtitling, only the component with subtitles created for Flemish television was selected. This subcorpus consists of more than 18 million words ( $n = 18,687,891$ ) and contains closed intralingual (i.e. source language is either Belgian Dutch or Netherlandic Dutch) and open interlingual (i.e. source language is English) subtitles that were compiled and broadcast by the Flemish public television station VRT<sup>5</sup> (with the channels *Eén* and *Canvas*) between 2000 and 2005. In order to investigate the dispersion of BSD vs. CBD in our corpus, we used the profile-based approach (Speelman, Grondelaers, & Geeraerts, 2003). This technique implies that the proportion of BSD features is studied in combination with the proportion of their CBD linguistic alternatives. Therefore, we compiled various lists of linguistic profiles, i.e. sets of language variants that cover the same meaning or linguistic function, with at least one variant being the BSD and one variant being the CBD. In other words, if the profile-based approach is applied, each variant in the profile should be able to replace its linguistic counterpart(s). All extracted data were manually validated to fulfill this condition. More advantages of the profile-based approach in the context of translation studies can be found in De Sutter et al. (2012) and Delaere and De Sutter (2013). In total, three sets of linguistic profiles were extracted from the SoNaR corpus, viz. a set with lexical-paradigmatic profiles (cf. Table 1), a set with constructional-paradigmatic profiles (cf. Table 2), and a set with syntagmatic profiles (cf. Table 3). Appendix 1<sup>6</sup> contains a representative selection of corpus examples of each of the profiles. Table 1 provides an overview of the lexical-paradigmatic profiles that were used for this study. The lexemes in each profile belong to the same lexical-semantic paradigm (i.e. they have the same denotation).

For the grammatical profiles, we made a distinction between constructional-paradigmatic and syntagmatic profiles. Table 2 shows the constructional-paradigmatic profiles, consisting of interchangeable constructions with the same meaning or function. Table 3 presents the syntagmatic profiles, which contain word order alternatives.

For the compilation of these lists, we consulted a number of normative sources that had to agree in characterizing the variants in each profile as BSD or CBD. The constructional-paradigmatic and syntagmatic profiles are based on the *VRT-Stijlboek* (*VRT Style Guide*)

(Hendrickx, 2003), whereas the lexical-paradigmatic profiles were extracted from the *Referentiebestand Belgisch-Nederlands (Reference File Belgian Dutch)* (Martin, 2005), a collection of 4,000 typical Belgian-Dutch words and expressions. In a last step, Van Dale's dictionary (Den Boon & Geeraerts, 2005; Geerts & den Boon, 1999) was consulted to verify the status of the lexical profiles. The data extraction resulted in a total of 36,518 validated observations (BSD:  $n = 27,109$ ; CBD:  $n = 9,409$ ). Tables 1–3 provide an overview of the number of attestations of each profile.

**Table 1.** Overview of the lexical-paradigmatic profiles used in this study.

Profile	BSD	CBD	Translation or meaning
1	<i>autosnelweg</i> ( $n = 33$ ) <i>autoweg</i> ( $n = 90$ ) <i>snelweg</i> ( $n = 129$ )	<i>autostrade</i> ( $n = 34$ )	<i>motorway</i>
2	<i>bestelwagen</i> ( $n = 99$ ) <i>bestelauto</i> ( $n = 3$ )	<i>camionette</i> ( $n = 234$ )	<i>delivery van</i>
3	<i>fiets</i> ( $n = 1257$ ) <i>rijwiel</i> ( $n = 7$ )	<i>velo</i> ( $n = 61$ )	<i>bicycle</i>
4	<i>handtas</i> ( $n = 186$ )	<i>sacoche</i> ( $n = 157$ )	<i>handbag</i>
5	<i>jas</i> ( $n = 643$ )	<i>frak</i> ( $n = 51$ )	<i>coat</i>
6	<i>krant</i> ( $n = 1026$ ) <i>dagblad</i> ( $n = 20$ )	<i>gazet</i> ( $n = 232$ )	<i>newspaper</i>
7	<i>laars</i> ( $n = 90$ )	<i>bot</i> ( $n = 24$ )	<i>boot</i>
8	<i>motor</i> ( $n = 392$ ) <i>motorfiets</i> ( $n = 43$ )	<i>moto</i> ( $n = 206$ )	<i>motorbike</i>
9	<i>nieuwsgierig</i> ( $n = 320$ ) <i>benieuwd</i> ( $n = 767$ )	<i>curieus</i> ( $n = 142$ )	<i>curious</i>
10	<i>oom</i> ( $n = 334$ ) <i>ome</i> ( $n = 8$ )	<i>nonkel</i> ( $n = 634$ )	<i>uncle</i>
11	<i>het platteland</i> ( $n = 147$ )	<i>de buiten</i> ( $n = 44$ )	<i>countryside</i>
12	<i>schrikken</i> ( $n = 894$ )	<i>verschieten</i> ( $n = 442$ )	<i>to be frightened</i>
13	<i>stropdas</i> ( $n = 12$ ) <i>das</i> ( $n = 148$ )	<i>plastron</i> ( $n = 31$ )	<i>tie</i>
14	<i>vrachtwagen</i> ( $n = 412$ )	<i>camion</i> ( $n = 127$ )	<i>truck</i>
15	<i>wastafel</i> ( $n = 14$ ) <i>wasbak</i> ( $n = 11$ )	<i>lavabo</i> ( $n = 49$ )	<i>sink</i>

**Table 2.** Overview of the constructional-paradigmatic profiles used in this study.

Profile	BSD	CBD	Translation or meaning
1	<i>adj + om + te + inf</i> ( $n = 225$ )	<i>adj + om + inf</i> ( $n = 34$ )	<i>adj + (to) + inf</i>
2	<i>akkoord gaan met</i> ( $n = 41$ )	<i>akkoord zijn met</i> ( $n = 7$ )	<i>to agree with</i>
3	<i>beginnen te + inf</i> ( $n = 244$ )	<i>beginnen + inf</i> ( $n = 460$ )	<i>to start (to) + inf</i>
4	<i>een beroep doen op</i> ( $n = 68$ )	<i>beroep doen op</i> ( $n = 32$ )	<i>to make an appeal to</i>
5	<i>durven te + inf</i> ( $n = 390$ )	<i>durven + inf</i> ( $n = 1355$ )	<i>to dare (to) + inf</i>
6	<i>mocht(en)</i> ( $n = 253$ ) <i>als</i> ( $n = 2413$ )	<i>moest(en)</i> ( $n = 257$ )	hypothetical clause
7	<i>niet hoeven</i> ( $n = 1060$ )	<i>niet moeten</i> ( $n = 1550$ )	<i>do not have to</i>
8	<i>op het eerste gezicht</i> ( $n = 137$ )	<i>op het eerste zicht</i> ( $n = 36$ )	<i>at first sight</i>
9	<i>over</i> ( $n = 91$ )	<i>na</i> ( $n = 69$ )	time indication
10	<i>passief</i> ( $n = 9575$ )	<i>passief + geworden/geweest</i> ( $n = 197$ )	<i>passive clause</i>
11	<i>prep + prep + en</i> ( $n = 222$ )	<i>prep + prep</i> ( $n = 549$ )	<i>sequence of prepositions</i>
12	<i>proberen te + inf</i> ( $n = 317$ )	<i>proberen + inf</i> ( $n = 8$ )	<i>to try (to) + inf</i>
13	<i>zeker weten dat</i> ( $n = 208$ ) <i>er zeker van zijn dat</i> ( $n = 122$ )	<i>zeker zijn dat</i> ( $n = 635$ )	<i>to be sure of</i>
14	<i>(zo)als + su</i> ( $n = 408$ )	<i>(zo)als + ob</i> ( $n = 6$ )	<i>like + object</i>
15	<i>zodra</i> ( $n = 483$ )	<i>van zodra</i> ( $n = 95$ )	<i>as soon as</i>
16	<i>zulke + plural noun</i> ( $n = 845$ )	<i>zo'n + plural noun</i> ( $n = 316$ )	<i>such + plural noun</i>

**Table 3.** Overview of the syntagmatic profiles used in this study.

Profile	BSD	CBD	Translation or meaning
<b>1</b>	<i>part + aux + inf</i> (n = 900) <i>aux + inf + part</i> (n = 349)	<i>aux + part + inf</i> (n = 689)	Position of the participle in the verbal end group
<b>2</b>	<i>NP + aux + inf</i> (n = 349)	<i>aux + NP + inf</i> (n = 10)	Position of the noun phrase in the verbal end group
<b>3</b>	<i>PA + aux + inf</i> (n = 1244)	<i>aux + PA + inf</i> (n = 573)	Position of the pronominal adverb in the verbal end group
<b>4</b>	<i>part + inf + inf</i> (n = 113)	<i>inf + part + inf</i> (n = 63)	Position of the infinitive in the verbal end group

### 3.3. The genres

In a next step, all data were manually annotated for the contextual parameters (*program*) *genre* and *source language* (the source language of the speaker was either English, Netherlandic Dutch or Belgian Dutch). As we wanted to apply a more fine-grained genre classification than Prieels et al. (2015), the annotation of (*program*) *genre* in this study was largely based on the genre division of Creeber (2008), who anecdotally outlined the repertory of genres in the television landscape. He distinguished 10 main genres, subdividing each of them into various subgenres (cf. appendix 2 for an overview of Creeber's classification, which will be discussed below). This classification, however, is not well defined and the genre definitions are therefore not always mutually exclusive. Consequently, this approach does not fit our purposes, since we want to categorize each television program into a unique genre and use this genre classification as a key factor in multivariate analysis. As a result, we had to adapt Creeber's classification to our specific research needs. To ensure that our research purposes would not influence the classification, we defined three contextual subparameters: *main purpose*, *target audience* and *cast*. These parameters allow us to identify and categorize different program genres. These parameters are comparable to what Biber & Conrad (2009) called the 'situational context', in which a certain register or language variety is used: "linguistic features tend to occur in a register because they are particularly well suited to the purposes and situational context of the register." (Biber & Conrad, 2009, p. 6). Table 4 provides an overview of these five program genres and their distinctive parameters (cf. appendix 3 for an explanation of the parameters and values in the table).

The left column in Table 4 contains the genres that were used to categorize the 109 television programs in our corpus. Compared with the genre division of Creeber (2008), our classification has been reduced to five program genres, since various genres (e.g. *costume drama*, *television news*, *music on television*) did not appear in the SoNaR corpus. In a last step, we applied the interannotator agreement procedure (Nowak & R ger, 2010) in order to validate our classification. Two independent annotators were asked to categorize the television programs in the corpus into different genres by using our classification table. This procedure had two phases: the first annotation round

**Table 4.** Overview of the genre classification used in this study.

Program genre	Main purpose	Target audience	Cast
<b>Fiction</b>	Entertaining	All ages	Actors
<b>Comedy</b>	Laughing	All ages	Actors
<b>Children's television</b>	Infotaining	Children	Actors + nonactors
<b>Light entertainment</b>	Infotaining	All ages	Nonactors
<b>Documentaries</b>	Informing	All ages	Nonactors

showed that some genres were not successfully defined, which resulted in a problematic classification of various TV programs. After having modified the values in our table, most of the television programs could be categorized successfully. Only a few programs (e.g. the weather report, the King's Christmas speech) did not fit our classification, so they were left out (cf. appendix 4 for a list of the television programs in each genre category). A brief description of the genres that were used in this study can be found below.

### 3.3.1. Fiction (drama)

Fiction programs are television programs of which the main purpose is to *entertain* the audience (the typical examples are *drama series* and *soaps*). These programs are intended for an audience of *all ages* (there is no particular age group) and the cast consists of *actors* playing a role by using a scripted text. *Fiction* is a collective name for what Creeber (2008) named *drama* (subdivided into *single play*, *western*, *action series*, *crime series*, *hospital drama*, *science fiction*, *mini-series*, *costume drama*, *teen series*, *drama-documentary* and *postmodern drama*) and *soap opera* (*soaps*). Given that our television corpus contains only 17 drama programs in total, we could not apply Creeber's fine-grained subgenre division as many of the subgenres would be left with no or only one program. This would have resulted in an unreliable empirical base for statistical analysis.

### 3.3.2. Comedy

The main purpose of comedy television programs is to make the audience *laugh* (the typical examples are *sketch* and *stand-up comedy*). These programs are also intended for an audience of *all ages* and the cast consists of *actors* playing a role by using a scripted text.

### 3.3.3. Children's television

The main purpose of children's television programs is to *infotain* the audience, which are mainly *children*. In other words, children's television wants to inform (or educate) children in an entertaining way. The cast in these programs consists of both *actors* (playing a role by using a script) and *nonactors* (dialogic text). This genre category also includes one cartoon (*De Avonturen van Kuifje*), because its episodes are often based on current affairs to inform the children about political and cultural events in an entertaining way.

### 3.3.4. Light entertainment

Light entertainment programs aim to inform the audience in an entertaining way, thus pursuing an *infotaining* purpose (the typical example is a *docusoap*). These programs are intended for an audience of *all ages* and the cast consists of both *actors* (playing a role by using a script) and *nonactors* (dialogic text). Typical programs within this genre are docusoaps, reality TV and game and talk shows. However, honesty compels us to admit that not every television program was easy to categorize and sometimes we had to make difficult decisions. A quiz show such as *Pappenheimers* differs a lot from a docusoap like *Het Leven Zoals Het Is*. However, both programs were categorized as *light entertainment*, since they are both intended for an audience of all ages and the cast consists of actors and nonactors. Furthermore, *Pappenheimers* and *Het Leven Zoals Het Is* each aim to entertain and inform the audience in their own way. *Pappenheimers* carries knowledge and facts during a game, whereas *Het Leven Zoals Het Is* informs the audience about a

particular profession (e.g. veterinarians) or working environment (e.g. hospital) by following these people during their daily practices.

### 3.3.5. Documentaries

The main purpose of documentaries is to *inform* the audience (a typical example is a reportage, introduced by a presenter and/or commented by a narrative voice). These programs are intended for an audience of *all ages* and the cast consists of both *actors* (playing a role by using a script) and *nonactors* (dialogic text). Within this genre, Creeber (2008) distinguished between *observational documentary* and *educational programming*. As Creeber mentioned ‘the provision of educational television has always targeted **children**’ (Creeber, 2008, p. 131; authors’ emphasis), we categorized the television programs of *educational programming* as *children’s television*. Consequently, the remaining subgenre *observational documentary* was named *documentaries*.

### 3.4. Statistical analysis: profile-based correspondence analysis

To verify how the language variants relate to the contextual parameters *source language* and *genre*, we applied profile-based correspondence analysis (Plevoets, 2015). This statistical technique analyzes the associations between the rows (*profiles*) and the columns (*contexts*) of a frequency table and visualizes these associations in a two-dimensional plot. Table 5 contains a section of the frequency table that was used for this analysis, displaying the total number of attestations for every language variant in each program genre.

In a first step, the correspondence analysis calculates two matrices with distances, one for the distances between the rows (e.g. the association between the variants *akkoord gaan met* and *akkoord zijn met* [to agree with] for the different program genres and source language varieties) and one for the distances between the columns (e.g. the association between the genre *fiction* and the genre *children’s television* for all language variants). This calculation is based on the chi-square test. Second, the resulting distances are visualized in a two-dimensional plot by reducing the original, multidimensional matrices to two-dimensional matrices. The distances in these two reduced matrices are then rendered in a biplot (i.e. a type of exploratory graph that generalizes the simple two-variable scatterplot), in which the distance between two data points indicates the exact nature of the association between them: the smaller the distance between the linguistic variants, the more closely they are related to each other (and vice versa). In other words, the relative

**Table 5.** Overview of the profile frequencies per program genre.

Variants	Label	Fiction	Children’s television	Light entertainment	Comedy	Documentaries
akkoord gaan met	BSD	27	4	3	2	5
akkoord zijn met	CBD	4	0	2	0	1
nieuwsgierig	BSD	121	57	91	5	46
benieuwd		266	63	397	29	12
curieus	CBD	123	1	11	3	4
part + aux + inf	BSD	218	72	246	21	343
aux + inf + part		49	6	56	14	224
aux + part + inf	CBD	249	34	209	37	160
zodra	BSD	87	20	257	11	108
van zodra	CBD	33	2	44	1	15
...	...	...	...	...	...	...

distances between the data points and the way they are clustered determine the interpretation of the results. In this study, we opted for the *profile-based* version of correspondence analysis, since this method has repeatedly been of great benefit in similar studies (cf. Speelman et al., 2003). An advantage of this profile-based approach is that the language variants are considered as subtypes within the same profile, rather than as isolated words. For instance, the lexemes *vrachtwagen*, *vrachtauto* and *camion* [truck] were treated as subtypes of the concept *vrachtwagen*. For more information about this technique and its advantages, we refer to De Sutter et al. (2012), Delaere et al. (2013) and Ghyselen (2015).

## 4. Results and discussion

In this section, the results of the profile-based correspondence analysis are presented and discussed. Section 4.1 provides some general information about the interpretation of the two-dimensional plot, which visualizes the linguistic choices made by Flemish subtitlers. In the following sections, the influence of the contextual parameters *source language* (4.2.) and *program genre* (4.3.) is discussed in more detail.

### 4.1. General observations

Table 6 visualizes the distribution of BSD and CBD in the subtitles on Flemish television. With regard to the lexical profiles, 74% of the variants are BSD, whereas 26% are CBD lexemes. The constructional profiles show more or less the same proportion: 75% of the variants are BSD constructions, whereas 25% are CBD. For the syntagmatic profiles, the number of CBD variants is somewhat higher (31%), resulting in a lower number of BSD variants (69%). These observations confirm our first hypothesis, viz. that subtitlers tend to avoid nonstandard lexicon. However, this table does not provide us with information about (i) the mutual behavior of the individual language variants and (ii) the contexts in which subtitlers opt for standard language or nonstandard language. Therefore, we applied profile-based correspondence analysis to visualize the linguistic choices of the subtitlers.

Figure 1 presents the distribution of the lexical, constructional and syntagmatic variants in the subtitles of Belgian-Dutch, Netherlandic-Dutch and English speakers on the one hand and five program genres on the other. The BSD variants are represented in gray and the CBD alternatives are marked in black. If we look at the dispersion of the linguistic variants, we can see that the first dimension (from left to right along the horizontal X axis) is defined by the dispersion of BSD vs. CBD, since the majority of the gray BSD variants are situated at the right side of the plot, whereas the black CBD variants are mainly located at the left side. Along the second dimension of this plot (from top to bottom along the vertical Y axis), we can observe genre- and source language-related variation, with

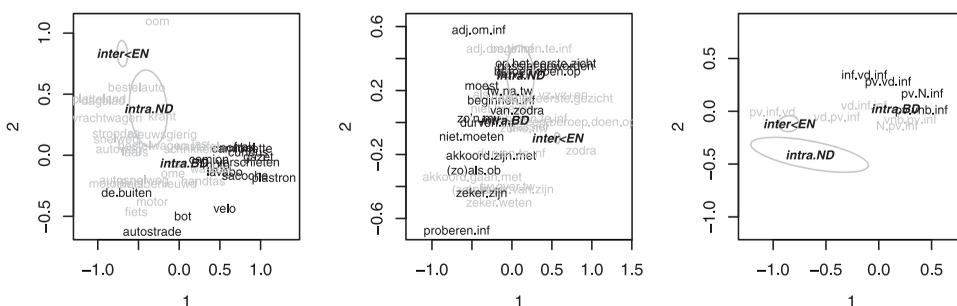
**Table 6.** Overview of the total number of BSD and CBD attestations per dataset

Dataset	BSD		CBD	
	Absolute	Relative (%)	Absolute	Relative (%)
Lexical profiles	7053	74	2468	26
Constructional profiles	17102	75	5606	25
Syntagmatic profiles	2955	69	1335	31



the contextual parameter *source language*. The source language is either Belgian Dutch (*intra.BD*), English (*inter < EN*) or Netherlandic Dutch (*intra.ND*). The different source language varieties are plotted upon the lexical, the constructional and the syntagmatic profiles (Figure 2).

The resulting biplots reveal some interesting findings. First, when we look at the position of the linguistic profiles, it can be observed that the variants are dispersed differently in the three plots. In the plot with the lexical profiles (left plot), most of the gray variants are located at the left side, whereas the black variants are clustered at the right side of the plot. This clearly indicates a strong divide between the contexts in which BSD variants are commonly and homogeneously used and those in which CBD variants are used, suggesting that subtitlers consciously choose lexical items depending on the context. With regard to the grammatical profiles, a comparable dispersion can be observed for the syntagmatic profiles, as the gray BSD variants are generally located at the bottom side of the plot and the black CBD variants are clustered at the upper side of the plot. The plot with the constructional profiles (middle plot), on the other hand, does not clearly divide the gray and the black variants, which implies that in all programs the subtitles contain both BSD and CBD. In other words, the source language of the program hardly has an influence on the use of standard and nonstandard constructional variants in the subtitles on Flemish television. This observation partially confirms our second hypothesis, viz. that the use of standard and nonstandard grammatical constructions is less context dependent compared to the use of standard and nonstandard lexemes. Second, when we look at the position of the different source language varieties relative to the dispersion of the linguistic variants, it can be observed that linguistic choices made in each of these contexts are in each dataset significantly different ( $p < .05$ ), since the ellipses do not overlap. It can clearly be seen that the interlingual subtitles of English television programs are located closest to the BSD variants in the three plots, which implies that in this context subtitlers use standard language to a large extent. However, this association with the BSD variants is less outspoken for the constructional profiles, since the distance from the *inter < EN* subtitles to the black CBD variants is smaller. Similarly, the subtitles of Netherlandic-Dutch programs are clearly related to the BSD variants for the lexical and syntagmatic profiles, whereas for the constructional profiles this source language variety is also surrounded by some CBD variants. Finally,



**Figure 2.** (from left to right) Biplot of the lexical-paradigmatic, constructional-paradigmatic and syntagmatic variants and the *source language* varieties (gray = BSD, black = CBD).

the intralingual subtitles of Belgian-Dutch programs contain a lot of nonstandard language, since this source language variety is located toward the CBD variants in the three plots. These findings verify hypothesis 3d and simultaneously confirm the results of our previous study (cf. Prieels et al., 2015). This lower level of norm-adherence in intralingual subtitles of Belgian-Dutch speakers can be explained by taking into account that in this context, subtitlers are directly exposed to original Belgian-Dutch speech. As already mentioned, CBD variants occur very frequently in spoken language on Flemish television. Consequently, it is to be expected that the subtitles of these programs also contain a high amount of CBD, because subtitlers plausibly transfer the CBD variants in the original Belgian-Dutch speech to the subtitles. In other words, the higher degree of nonstandard language in intralingual subtitles of Flemish speakers is most likely caused by direct interference of the language use in the original Belgian-Dutch television program. However, we were not able to consult the original footage of the television programs, since the SoNaR corpus does not contain the spoken source texts. Further research, which is currently being undertaken, will aim to substantiate these assumptions.

#### 4.3. The influence of (program) genre on the subtitlers' linguistic choices

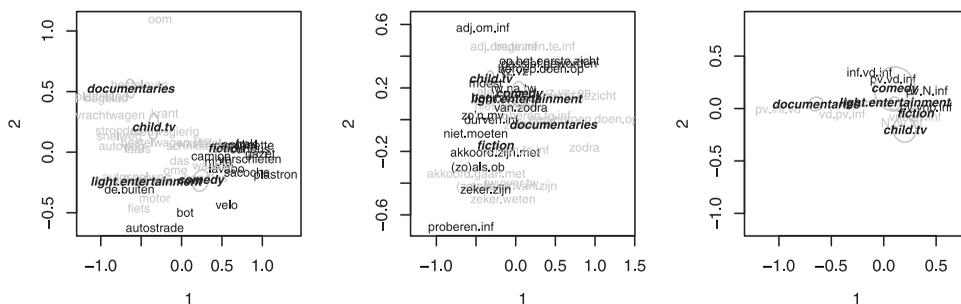
Further analyses revealed not only that the source language influences the subtitlers' linguistic choices, but that the contextual parameter *program genre* is also a determining factor. Especially the intralingual subtitles of Belgian-Dutch television programs (and, to a lesser extent, the intralingual subtitles of Netherlandic-Dutch television programs) show significant genre variation. The program genre does not influence the linguistic behavior of the interlingual subtitles of English programs, since these subtitles mainly contain standard language (cf. section 4.2.). In this section, we discuss the influence of the program genre on the linguistic choices of the subtitlers in more detail.

In general, the linguistic choices made in each of the genre varieties are significantly different ( $p < .05$ ) in the three datasets, since the ellipses do not overlap (except for the genres *light entertainment* and *comedy* in the syntagmatic plot). Further, the position of the different genres relative to the dispersion of the linguistic variants shows that there is a lot of genre variation, which is particularly caused by the influence of the Belgian-Dutch spoken programs in our dataset. The English and Netherlandic-Dutch spoken programs show (almost) no variation (cf. supra). First, in the plot with the lexical profiles (left plot), it can clearly be seen that *fiction* is the genre that is most related to the CBD. On the contrary, subtitles in *comedy* and *light entertainment* are more related to the BSD variants, although these genres are still surrounded by some of the CBD variants, whereas the subtitles in *documentaries* and *children's television* are mostly related to the BSD variants, with their distance to the nonstandard variants being larger. Second, in the plot with the constructional profiles (middle plot) the program genre hardly has an influence on the use of standard and nonstandard constructions in the subtitles on Flemish television. This observation confirms once again that the use of standard and nonstandard grammatical constructions is less context dependent (cf. hypothesis 2). Finally, in the plot with the syntagmatic profiles (right plot) the genre *comedy* is located closest to the CBD variants, whereas *light entertainment* and *fiction* are located closer to the BSD variants. *Documentaries* and *children's television*, on the contrary, are the most norm-adhering genres, since the distance from these genres to CBD variants is the largest. Furthermore, if we compare

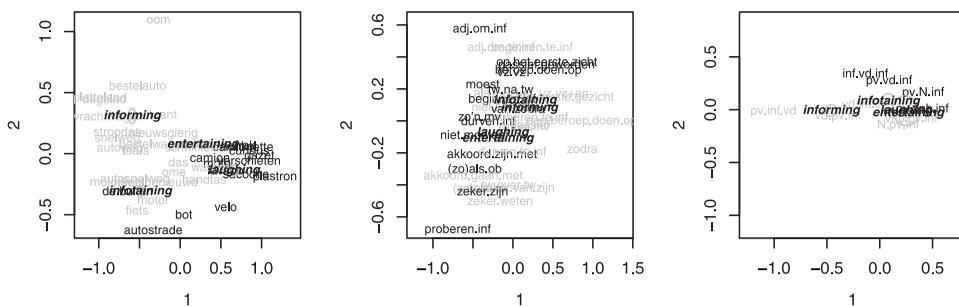
the three datasets, we can notice that, on the one hand, the relative distance of *children's television* to the CBD variants is smaller for the constructional and syntagmatic profiles than for the lexical profiles. On the other hand, the relative distance of *light entertainment* to the core of the CBD variants is larger for the lexical profiles than for the constructional and syntagmatic profiles. The main conclusions that can be drawn from these findings are the following. First, the results demonstrated that the contextual parameter *program genre* affects the linguistic choices of subtitlers. The analyses revealed that subtitles in *documentaries* and *children's television* mainly contain standard language, whereas subtitles in *fiction* and *comedy* contain a lot of nonstandard language and subtitles in *light entertainment* take a middle position (verification hypothesis 3). Simultaneously, it was shown that subtitlers more often avoid nonstandard lexemes than nonstandard constructional and syntagmatic variants in certain genres (cf. appendix 5 for the total number of BSD and CBD attestations per dataset in each genre). This conclusion confirms our first hypothesis, viz. that Flemish subtitlers are more norm-adhering toward lexical variants than toward constructional and syntagmatic variants. A possible explanation for this outcome is that subtitlers are more aware of standard and nonstandard lexemes, since these features are more salient than nonstandard constructions are (cf. Lybaert, 2014). Furthermore, the subtitle guidelines of the Flemish public broadcaster give the impression that the sensitivity toward the use of standard and nonstandard lexemes is remarkably higher. From this norm-adherent point of view of VRT, subtitlers plausibly aim to avoid nonstandard lexemes in certain program genres. So far, the analyses have revealed in which genres subtitlers tend to or tend not to use BSD. However, we want to investigate which specific features of these genres (*program purpose*, *target audience* and *cast*) determine the subtitlers' linguistic choices. What we still do not know, for instance, is (i) whether subtitles contain more nonstandard constructional variants than nonstandard lexical and syntagmatic variants in entertaining programs and (ii) whether actors are subtitled differently in comparison with nonactors. To answer these questions, we calculated and visualized the relative distances between the subparameters and the linguistic profiles.

#### 4.3.1. The influence of program purpose on the subtitlers' linguistic choices

Figure 3 is basically the same as Figure 4, but here it represents the position of the different *program purposes* relative to the position of the linguistic variants. With regard to the



**Figure 3.** (from left to right) Biplot of the lexical-paradigmatic, constructional-paradigmatic and syntagmatic variants and the *program purpose* varieties (gray = BSD, black = CBD).



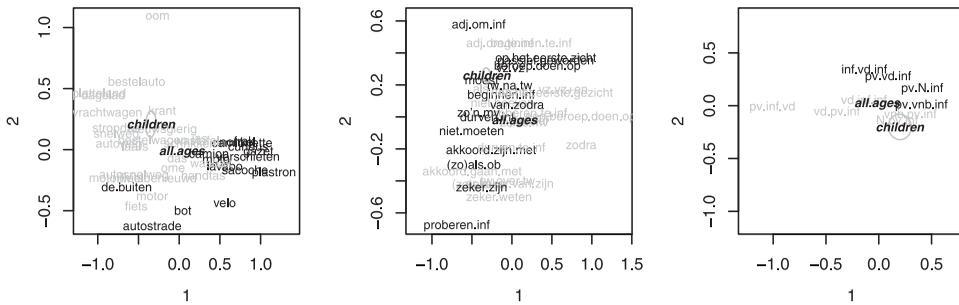
**Figure 4.** (from left to right) Biplot of the lexical-paradigmatic, constructional-paradigmatic and syntagmatic variants and the *program genre* varieties (gray = BSD, black = CBD).

lexical profiles (left plot), the subtitles' linguistic choices are significantly different in the four subtitle contexts. It can immediately be observed that subtitles of *laughing* programs are most clearly related to CBD variants, whereas subtitles of *informing* programs are most related to BSD variants. Subtitles of *infotaining* and *entertaining* programs are also related to the BSD variants, although less notable than informing programs, as they are still surrounded by some CBD variants. As mentioned above, the use of standard language vs. nonstandard language is less context dependent for the constructional profiles (middle plot) than for the lexical (left plot) and syntagmatic (right plot) profiles, since the position of the constructional CBD and BSD variants is extremely varied. Based on the ellipses, it can be seen that the linguistic choices of the subtitles in *informing* and *infotaining* programs, on the one hand, and *laughing* and *entertaining* programs, on the other, are similar for the constructional profiles. In the plot with the syntagmatic variants, the *program purpose* affects the linguistic behavior of the subtitles less, since the variation between the different contexts is less prominent. It can be observed that subtitles in *informing* programs contain more BSD variants than do subtitles in *infotaining*, *entertaining* and *laughing* programs, which are located much closer to the CBD variants.

In sum, the analyses showed again that in some television programs subtitles more often avoid the use of nonstandard lexemes than nonstandard constructions (verification of hypothesis 1). Furthermore, the results demonstrated that the program purpose influences the linguistic choices of subtitles. The lower degree of norm-adherence in noninformative television programs (entertaining and laughing) can be explained by their general aim. The main objective of entertaining and humorous programs is to amuse the audience, thus creating an informal, spontaneous atmosphere, which has a greater chance of showing spontaneous, colloquial utterances. As a consequence, it seems plausible that subtitles want to reproduce this spontaneous nature of the television program by using nonstandard, colloquial features in the subtitles (verification of hypothesis 3a).

#### 4.3.2. The influence of target audience on the subtitles' linguistic choices

In Figure 5, the influence of the *target audience* is visualized. The most interesting observation is that subtitles in television programs intended for *children* are clearly related to the BSD variants in the lexical (left plot) and the syntagmatic plot (right plot), whereas these subtitles contain a lot of nonstandard constructional variants (middle plot). Subtitles in



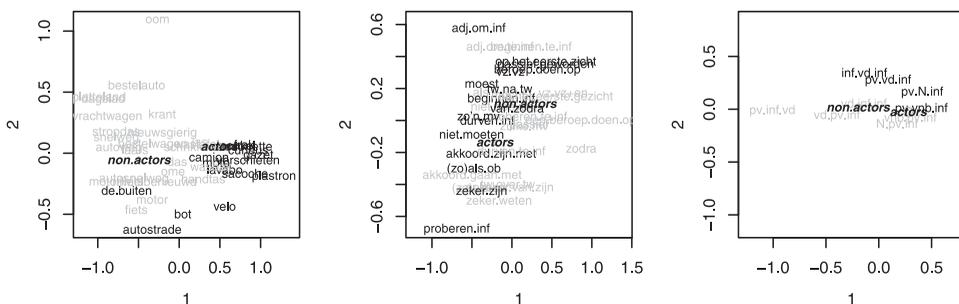
**Figure 5.** (from left to right) Biplot of the lexical-paradigmatic, constructional-paradigmatic and syntagmatic variants and the *target audience* varieties (gray = BSD, black = CBD).

television programs without a particular age group (*all ages*) are generally more related to the CBD variants, although their closer position is less notable for the syntagmatic and constructional profiles. In the lexical plot, however, the relative distance of *children* to the CBD lexemes is larger than the relative distance of *all ages* to the CBD lexemes, which implies that the subtitles of children's television programs particularly contain standard language at lexical level.

The main conclusion emerging from Figure 5 is that the influence of the *target audience* on the linguistic behavior of subtitles is generally less explicit, except for the lexical profiles. The analyses demonstrated that subtitlers more often tend to avoid nonstandard lexemes in television programs intended for children (cf. appendix 5 for the total number of BSD and CBD attestations per dataset for *target audience*). Furthermore, subtitles in children's television programs generally contain more standard language than subtitles in programs without a particular age group (verification of hypothesis 3b). This could be explained by the educational footing of children's programs. These programs aim to perform an exemplary role, also on the level of language use, which accounts for the recurrent use of standard language, both in the spoken language and in the subtitles.

#### 4.3.3. The influence of cast on the subtitlers' linguistic choices

Figure 6 visualizes the influence of *cast* on the lexical (left plot), the constructional (middle plot) and the syntagmatic (right plot) profiles. The three plots show that the linguistic



**Figure 6.** (from left to right) Biplot of the lexical-paradigmatic, constructional-paradigmatic and syntagmatic variants and the *cast* varieties (gray = BSD, black = CBD).

choices made in subtitles of *actors* and *nonactors* are significantly different, as the ellipses do not overlap. Subtitles of *actors* are most related to the CBD variants in both the lexical and syntagmatic plot, whereas subtitles of *nonactors* are located closer to the BSD variants. Especially in the lexical plot, the relative distance of *nonactors* to the CBD lexemes is larger than the relative distance of *actors* to the CBD lexemes, which implies that subtitles of *nonactors*, in particular, contain standard language at the lexical level. For the constructional profiles, however, subtitles of *actors* and *nonactors* both contain standard and nonstandard language, since they are surrounded by BSD variants as well as by CBD variants.

In other words, the analyses revealed that *cast* is an important factor in the linguistic choices of subtitlers. Contrary to what was assumed in hypothesis 3c, subtitles of *nonactors* tend to contain standard language, especially for the lexicon, whereas subtitles of *actors* contain a lot of nonstandard language. A similar ‘linguistic hierarchy’ was observed in Remael et al. (2008). In their study, it was demonstrated that none of the television hosts were subtitled, whereas all of the interviewees were. This implies that, unlike television hosts, the language use of interviewees is expected to contain a lot of nonstandard language and be unintelligible for the viewers and, as a consequence, needs to be ‘translated into Standard Dutch’ in the subtitles. This could explain why the subtitles of *nonactors* (e.g. interviewees) contain more BSD than those of *actors* (who are also involved in the making of the program). Furthermore, Remael (2003, p.226) emphasized that ‘it is important to distinguish the scripted dialogue (of *actors* [authors’ addition]) of fiction films or TV series from the more or less spontaneous speech of a live interview (*nonactors* [authors’ addition])’, because ‘in fictional dialogue both register and the interactional features of conversation are part of a carefully constructed narrative that also relies on other sign systems to communicate with the viewer’. In other words, the use of nonstandard language in the subtitles of *actors* could be a conscious strategy in the communication system of the subtitlers. After all, the public broadcaster is more tolerant toward the use of nonstandard language to maintain the authenticity of the characters in entertainment programs, so it can be expected that the oral features in the original speech of the actors are transferred to the subtitles. In that way, the subtitlers want to avoid a situation where ‘the characters speak like a printed page’ (Rosa, 2001, p. 216). Furthermore, it was again demonstrated that subtitlers more often avoid nonstandard lexemes in subtitles of *nonactors* (cf. appendix 5 for the total number of BSD and CBD attestations per dataset for *cast*).

## 5. Conclusions

The aim of this study was to demonstrate how Flemish subtitlers deal with the linguistic tension between the norm-adherent language policy of the public broadcaster VRT, which is oriented toward the use of BSD, and the specific language reality (i.e. Flanders and its repertory of nonstandard varieties). Building on a corpus of subtitles that were produced by the Flemish public broadcaster between 2000 and 2005, we statistically analyzed the linguistic choices of subtitlers in various contexts. More specifically, we applied profile-based correspondence analysis to measure which contextual factors (program genre and source language) determine whether Flemish subtitlers opt for BSD or CBD lexemes and constructions. Our results demonstrated that source language of the speaker and program genre are two determining factors that cause norm-related differences in the subtitle corpus. First, subtitlers are more norm-adhering when

subtitling English or Netherlandic-Dutch spoken television programs compared to Belgian-Dutch spoken programs. In other words, the subtitles contained significantly more CBD variants when the source language of the original footage is (Colloquial) Belgian-Dutch. Second, the language use in the intralingual subtitles in Belgian-Dutch spoken programs was influenced by the program genre to a large extent. The subtitles in more informative programs (documentaries) and programs intended for children (children's television) contain standard language to a large extent, whereas the number of CBD lexemes and constructions increased significantly in the subtitles of humorous programs (comedy). Light entertainment and fiction take a middle position. A plausible explanation for these results is that subtitlers directly transfer the CBD variants in the original footage to the subtitles (cf. Prieels et al., 2015). Not only do these CBD variants frequently show up in Flemish speech, but the use of nonstandard language in television dialogue can be a conscious strategy to entertain the audience or to create a comic effect (e.g. McIlvenny, Mettovaara, & Tapio, 1992; Remael, 2003; Rutter, 1997). As a consequence, it should not surprise us that subtitlers want to maintain this 'linguistic effect' of the original footage in the subtitles. However, since we were not able to analyze the original, spoken television fragments, further research needs to be conducted to verify the aforementioned assumptions. Nevertheless, these results emphasize the importance of contextual factors that should be taken into account when analyzing subtitlers' linguistic choices, since our study pointed out that this linguistic behavior is largely oriented toward the aim of the television program and toward the target audience. These observations raise some new questions about the comprehensibility and suitability of nonstandard language in subtitling from the audience's perspective. In-depth research into the perception and attitude of the viewer toward norm-related language differences in subtitles could reveal some interesting findings for the subtitling field and audiovisual translation as a whole. Furthermore, the results showed that, in general, Flemish subtitlers are more norm-adhering toward CBD lexemes than toward CBD grammatical constructions. In other words, nonstandard lexemes occur less frequently than nonstandard grammatical constructions in subtitles on Flemish television. The most obvious explanation offered for these results is that CBD lexical features are significantly more often perceived by language users than CBD syntactic features are (Lybaert, 2014). This salience effect could explain why subtitlers tend to avoid nonstandard lexicon. Further, in-depth research into salience differences between language variants could provide a more fine-grained insight into how salience affects the subtitlers' linguistic choices.

## Notes

1. Website of the Dutch Language Union: <http://taaladvies.net>
2. VRT stands for *Vlaamse Radio- en Televisieomroeporganisatie*, the Flemish public broadcaster for radio and television in the Dutch-speaking part of Belgium, Flanders. Website: [www.vrt.be](http://www.vrt.be)
3. With regard to open interlingual subtitles, VRT aims for the use of standard language.
4. The Dutch Parallel Corpus is stratified across five genres: literature, journalistic texts, administrative texts, instructive texts and texts for external communication.
5. VRT confirmed that all subtitlers were Flemish by mother tongue.
6. The appendices can be consulted in a separate online document.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Notes on contributors

**Lynn Prieels** is a PhD student at the Department of Translation, Interpreting and Communication at Ghent University (Belgium). In October 2014, she was awarded a PhD fellowship by the Flemish Research Foundation (FWO) for a period of four years. Her current research focuses on linguistic norm-adherence in audiovisual translation. More specifically, she investigates lexical and syntactic variation in interlingual and intralingual Belgian Dutch subtitling. In the autumn of 2015 her first article on linguistic variation in Belgian Dutch subtitles was published.

**Gert De Sutter** is associate professor in translation studies and Dutch linguistics in the Department of Translation, Interpreting and Communication at Ghent University (Belgium). His main research interests involve the overlapping fields of quantitative corpus linguistics and translation studies. Research topics he is currently dealing with are normalization, conservatism, interference and risk aversion in corpus-based translation studies, as well as a variety of syntax-oriented topics in Dutch language studies (word order phenomena in Dutch).

## References

- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.
- Cavalheiro, L. (2008). Linguistic variation in subtitling for audiovisual media in Portugal. *Linguistica Antverpiensia*, 7, 17–28.
- Creeber, G. (2008). *The television genre book* (2nd ed.). London: Palgrave Macmillan.
- De Ridder, R. (2015). 'Arme rijke taal' *Audiovisual translation and minority language planning in pluricentric language areas: A case study of Flemish public service broadcasting (Doctoral dissertation)*. Dublin City University.
- De Sutter, G., Delaere, I., & Plevvoets, K. (2012). Lexical lectometry in corpus-based translation studies. Combining profile-based correspondence analysis and logistic regression modeling. In M. Oakes & M. Ji (Eds.) *Quantitative methods in corpus-based translation studies. A practical guide to descriptive translation research* 51 (pp. 325–346). Amsterdam/Philadelphia: John Benjamins.
- Delaere, I., & De Sutter, G. (2013). Applying a multidimensional, register-sensitive approach to visualize normalization in translated and non-translated Dutch. In M. A. Lefer, & S. Voegelers (Eds.), *Interference and Normalisation in Genre controlled Multilingual Corpora. Belgian Journal of Linguistics*, 27, 43–60. doi:10.1075/bjl.27.03del
- Delaere, I., De Sutter, G., & Plevvoets, K. (2012). Is translated language more standardized than non-translated language?: Using profile-based correspondence analysis for measuring linguistic distances between language varieties. *Target*, 24(2), 203–224. doi:10.1075/target.24.2.01del
- Den Boon, T., & Geeraerts, D. (2005). *Van Dale: Groot Woordenboek der Nederlandse Taal: 3 Dl.* [Van Dale: Dictionary Dutch Language: Vols 1-3] (14th ed.). Utrecht: Van Dale Lexicografie.
- Dewulf, B., & Saerens, G. (2000). *Stijlboek Teletekstondertiteling* [Style guide Teletext subtitling], internal document VRT-teletekstredactie.
- Diaz-Cintas, J. (2010). Subtitling. In Y. Gambier, & L.V. Doorslaer (Eds.), *Handbook of translation studies* (pp. 344–349). Amsterdam/Philadelphia: John Benjamins.
- Doens, E. (2000). *Een kwalitatief onderzoek van ondertiteling voor doven en slechthorenden* [A qualitative study on subtitling for the Deaf and Hard-of Hearing], internal document VRT studiedienst.
- Geeraerts, D. (1998). VRT-Nederlands en Soap-Vlaams [VRT-Dutch and Soap-Flemish]. *Nederlands van Nu*, 3, 75–77.
- Geeraerts, D. (2001). Een zondagspak? Het Nederlands in Vlaanderen: gedrag, beleid, attitudes. *Ons erfdeel*, 44(3), 337–343.

- Geeraerts, D., Grondelaers, S., & Speelman, D. (1999). *Convergentie en divergentie in de Nederlandse woordenschat: een onderzoek naar kleding- en voetbaltermen* [Convergence and divergence in the Dutch vocabulary; a study on clothing and football terms]. PJ Meertens-Instituut.
- Geerts, G., & den Boon C.A. (1999). *Van Dale: Groot Woordenboek der Nederlandse Taal: 3 Dl.* [Van Dale: Dictionary Dutch Language: Vols 1-3] (13th ed.). Utrecht: Van Dale Lexicografie.
- Ghyselen, A.-S. (2015). 'Stabilisering' van tussentaal? [Stabilization of intermediate language?]. *Taal en Tongval*, 67(1), 43–95. doi:10.5117/TET2015.1.GHYS
- Goossens, J. (2000). De toekomst van het Nederlands in Vlaanderen [The future of Dutch in Flanders]. *Ons Erfdeel*, 43(1), 3–13.
- Grondelaers, S., & Van Hout, R. (2011). The standard language situation in the low countries: Top-down and bottom-up variations on a diaglossic theme. *Journal of Germanic Linguistics*, 23(03), 199–243. doi:10.1017/S1470542711000110
- Hamaida, L. (2007). Subtitling slang and dialect. *Proceedings of the MuTra 2007-LSP translation scenarios conference*, 1–11.
- Hendrickx, R. (1998). Het taalcharter [The Language Charter]. VRTtaal.net. Retrieved from <http://www.vrt.be/taal/taalcharter>
- Hendrickx, R. (2003). *Stijlboek VRT* [Style guide VRT]. Utrecht: Van Dale Lexicografie.
- Janssens, G., & Marynissen A. (2008). *Het Nederlands vroeger en nu* [Dutch in the past and present]. (3rd ed.). Leuven/Voorburg: Acco.
- Karamitroglou, F. (2000). *Towards a methodology for the investigation of norms in audiovisual translation*. Amsterdam: Rodopi.
- Lefevere, E. (2011). *Tussentaal in de Vlaamse media* [Intermediate language in the Flemish media] (*Unpublished Master's Dissertation*). Ghent University.
- Lybaert, C. (2014). Perceptie van tussentaal in het gesproken Nederlands in Vlaanderen [Perception of intermediate language in spoken Dutch in Flanders]. *Nederlandse Taalkunde*, 19(2), 185–220. doi:10.5117/NEDTAA2014.2.LYBA
- Martin, W. (2005). *Het Belgisch-Nederlands anders bekeken: het Referentiebestand Belgisch-Nederlands (RBBN)* [Belgian Dutch from a different point of view: Reference File for Belgian Dutch]. Technical report. Amsterdam: Vrije Universiteit.
- McIlvenny, P., Mettovaara, S., & Tapio, R. (1992, May). I really wanna make you laugh: Stand-up comedy and audience response. *Folia, fennistica and linguistica: Proceedings of the annual Finnish linguistic symposium May*, 225–247.
- Neves, J. (2004). Language awareness through training in subtitling. In P. Orero (Ed.), *Topics in audiovisual translation* (pp. 127–140). Amsterdam & Philadelphia: Benjamins.
- Nowak, S., & Rüger, S. (2010). How reliable are annotations via crowdsourcing: A study about inter-annotator agreement for multi-label image annotation. *Proceedings of the international conference on multimedia information retrieval*, 557–566.
- Pinto, S. R. (2009). How important is the way you say it?: A discussion on the translation of linguistic varieties. *Target*, 21(2), 289–307. doi:10.1075/target.21.2.04pin
- 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* [Between colloquial and standard language: a corpus-based study in the situational, regional and social dispersion of morfosyntactic features in spoken Belgian Dutch] (*Unpublished dissertation*). Catholic University, Leuven.
- Plevoets, K. (2015). Corregp: Functions and Methods for Correspondence Regression. Ghent University. Retrieved from <https://cran.rproject.org/web/packages/corregp/index.html>
- Prieels, L. (2013). Tussentaal in tv-programma's [Intermediate language in television programs]. In J. De Caluwe, S. Delarue, A.-S. Ghyselen, & C. Lybaert (Eds.), *Tussentaal: over de talige ruimte tussen dialect en standaardtaal in Vlaanderen* (pp. 35–49). Gent: Academia Press.
- Prieels, L., Delaere, I., Plevoets, K., & De Sutter, G. (2015). A corpus-based multivariate analysis of linguistic norm-adherence in audiovisual and written translation. *Across Languages and Cultures*, 16(2), 209–231. doi:10.1556/084.2015.16.2.4

- Remael, A. (2003). Mainstream narrative film dialogue and subtitling. *The Translator*, 9(2), 225–247. doi:10.1080/13556509.2003.10799155
- Remael, A., De Houwer, A., & Vandekerckhove, R. (2008). Intralingual open subtitling in Flanders: Audiovisual translation, linguistic variation and audience needs. *Journal of Specialized Translation*, 10, 76–105.
- Reynaert, M., Oostdijk, N., De Clercq, O., van den Heuvel, H., & de Jong, F. (2010). Balancing SoNaR: IPR versus processing issues in a 500-million-word written Dutch reference corpus. *Seventh conference on international language resources and evaluation*, LREC '10, 2693–2698, 19–21 May 2010, Malta.
- Rosa, A. A. (1994). The centre and the edges. Linguistic Variation and Subtitling Pygmalion into Portuguese. *Translation and the (Re) Location of Meaning. Selected Papers of the CETRA Research Seminars in Translation Studies*, 317–338.
- Rosa, A. A. (2001). Features of oral and written communication in subtitling. In Y. Gambier, & H. Gottlieb (Eds.), *(Multi)media translation: Concepts, practices and research* (pp. 213–222). Amsterdam/Philadelphia: Benjamins Translation Library, 34.
- Rutter, J. (1997). *Stand-up as interaction: Performance and audience in comedy venues (Unpublished dissertation)*. University of Salford.
- Slembrouck, S., & Van Herreweghe, M. (2004). Teletekstondertiteling en tussentaal: De pragmatiek van het alledaagse [Teletext subtitling and intermediate language: Everyday pragmatics]. In J. De Caluwe, G. De Schutter, M. Devos, & J. Van Keymeulen (Eds.), *Taeldeman, man van de taal, schatbewaarder van de taal* (pp. 853–873). Ghent: Academia Press.
- Speelman, D., Grondelaers, S., & Geeraerts, D. (2003). Profile-based linguistic uniformity as a generic method for comparing language varieties. *Computers and the Humanities*, 37(3), 317–337. doi:10.1023/A:1025019216574
- Van Bree, C. (2000). Taalbewustzijn, taalverandering en regionale variatie. Een benadering [Linguistic awareness, language change and regional variation. An approach]. *Taal en Tongval*, 52(1), 22–46.
- Van Hoof, S. (2010). Taalvariatie op de Vlaamse openbare omroep: Substandaardisering in fictieprogramma's vroeger en nu [Language variation on the public broadcaster: Substandardization in fiction programs from the past and present] (*Unpublished dissertation*). Antwerp University.
- Van Hoof, S., & Vandekerckhove, B. (2013). Feiten en fictie. Taalvariatie in Vlaamse televisiereeksen vroeger en nu [Facts and fiction. Language variation in Flemish television series from the past and present]. *Nederlandse Taalkunde*, 18(1), 35–64. doi:10.5117/NEDTAA2013.1.HOOF
- Vandekerckhove, R., De Houwer, A., & Remael, A. (2007). Intralinguale ondertiteling op de Vlaamse televisie: een spiegel voor de taalverhoudingen in Vlaanderen? [Intralingual subtitling on Flemish television: A mirror for language balance in Flanders?]. In S. Dominiek et al. (Ed.) *Tussen taal, spelling en onderwijs. Essays bij het emiraat van Frans Daems*, (pp. 71–83). Gent: Academia Press.
- Vandekerckhove, R., De Houwer, A., & Remael, A. (2009). Between language policy and linguistic reality: Intralingual subtitling on Flemish television. *Pragmatics*, 19(4), 609–628. doi:10.1075/prag.19.4.05van
- Vandekerckhove, R., De Houwer, A., Remael, A., & Van der Niepen, I. (2006). Intralinguale ondertiteling van Nederlandstalige televisieprogramma's in Vlaanderen: linguïstische en extralinguïstische determinanten [Intralingual subtitling in Dutch-spoken television programs in Flanders: Linguistic and extra-linguistic determinants]. In T. Koole, J. Nortier, & B. Tahitu (Eds.), *Artikelen van de vijfde sociolinguïstische conferentie (Anéla - Nederlandse vereniging voor toegepaste taalwetenschap)* (pp. 503–513). Delft: Eburon.
- Willemyns, R., & Daniëls, W. (2003). *Het verhaal van het Vlaams: de geschiedenis van het Nederlands in de Zuidelijke Nederlanden*. Standaard.