

A mixed-method approach to the use of Colloquial Belgian Dutch in intralingual subtitling on Flemish television

Further evidence for the gradual acceptance of tussentaal

Lynn Prieels and Gert De Sutter¹

TET 70 (2): 211–256

DOI: 10.5117/TET2018.2.PRIE

Abstract

This paper addresses the question to what extent Flemish subtitlers reproduce Belgian-Dutch colloquialisms of the spoken source text in the subtitles. Next to the official Belgian Standard Dutch language, Colloquial Belgian Dutch or *tussentaal* is frequently spoken on Flemish television. In this context, it is particularly interesting to investigate whether the subtitlers copy those spoken colloquialisms to the subtitles or whether they translate them into standard language. This study compares the language used in twenty television programs to the corresponding subtitles in order to verify the linguistic choices of the subtitlers. In addition, it will be examined whether the subtitles contain more lexical colloquialisms than morphological or syntactic colloquialisms, and whether the program genre influences these choices. The results reveal that Colloquial Belgian Dutch lexemes are more often reproduced in subtitles on Flemish television than morphological and syntactic colloquialisms. Furthermore, it is shown that especially the subtitles of entertainment and comedy programs contain *tussentaal*. Based on these results, we conclude that the demotization process in Flanders is not confining itself to the spoken registers, since it also manifests itself in certain written contexts.

Keywords: audiovisual translation, subtitling, Belgian-Dutch, corpus-based, standardization, destandardization, demotization, language variation

1 Introduction

The last few decades, many Dutch linguists have focused on the repertory of Dutch language varieties (e.g. Belgian Standard Dutch, Colloquial Belgian Dutch, regiolect, dialect) that are used on Flemish television. Van Gijssel et al. (2008), for example, demonstrated that *tussentaal* (i.e. an intermediate, substandard register of Belgian Dutch, also called Colloquial Belgian Dutch) has taken up a central position in Flemish (radio and) TV commercials. In several other studies it was shown that *tussentaal* also appears in soaps, comedy programs and TV shows (e.g. De Ridder, 2007; Lefevere, 2011; Prieels, 2013; Van Hoof, 2013). Furthermore, a corpus-based analysis of the language used in the reality series *Expeditie Robinson* of Zenner et al. (2007) revealed that the use of this Colloquial Belgian Dutch variety (versus Belgian Standard Dutch) depends on the conversational situation in which Flemish speakers are involved. Given the growing use of Colloquial Belgian Dutch in spoken public media, the question arises whether this informal variety does also penetrate the written registers, which would (at least partly) undermine the current belief that Belgian Standard Dutch is the only appropriate variety to use in written language. Although the aforementioned studies offer a valuable insight into the spoken language use on Flemish television, research into Belgian-Dutch subtitling is still relatively scarce (e.g. De Ridder, 2015; Remael et al., 2008). Nevertheless, it would be exceedingly interesting to investigate how Flemish subtitlers deal with this language variation in the spoken source text, since Flemish subtitlers have to choose whether they reproduce the colloquial variants of the spoken source text rather than converting these colloquialisms into standard language. Previous research has already focused on language variation in subtitling, yielding many interesting insights. Cavalheiro (2008), for instance, showed that the substandard variety spoken in the film *Gone with the Wind* was translated into an ‘equivalent’ substandard Portuguese variety in the subtitles on the private television channel. Other studies, however, have demonstrated that non-standard language varieties (such as dialect, slang, regiolect) in the spoken source text are generally standardized (i.e. translated into standard language) in the corresponding subtitles. Not only are these non-standard linguistic features often difficult to reproduce in written language, subtitlers are also frequently bound by the language policy of the TV channels, which generally support the use of standard language (e.g. Hamaida 2007; Pinto, 2009; Remael et al. 2008; Rosa 2001). We want to verify whether this is also the case in subtitling on Flemish television.

In a geographical-oriented study, Vandekerckhove et al. (2006/2007) compared the intralingual subtitling practice of the Flemish public broadcasting company VRT and the commercial channel VTM. Based on a corpus of 793 fiction programs, it was shown that the Western regiolect was subtitled more often than the (dominant) Brabant regiolect on Flemish television. However, little to no attention went to the actual lexical, morphological, syntactic, and pragmatical variation within the subtitles and, as a consequence, these findings do not answer the question as to whether subtitlers make more use of Colloquial Belgian Dutch or Belgian Standard Dutch. In a recent study, De Ridder (2015) analyzed the use of Belgian-Dutch lexis in the subtitles of crime fiction series on the Flemish public broadcaster VRT. She found that intralingual subtitles (i.e. the source language is Dutch) contain significantly more Colloquial Belgian Dutch lexemes than interlingual subtitles (i.e. the source language is a foreign language, e.g. English). Nonetheless, this analysis merely focused on Colloquial Belgian Dutch lexicon (thereby omitting colloquial syntactic and morphological constructions), and it did not investigate the Flemish spoken source text and its potential influence on the subtitlers' linguistic choices. Furthermore, in the aforementioned studies the effect of extra-linguistic contexts (e.g. program genre) was not examined, since there was only one television genre taken into consideration.

To fill this gap, the present study compares the use of standard and colloquial lexical, syntactic, and morphological items in Belgian-Dutch subtitles to the corresponding Flemish spoken source text, and is based on a corpus of five program genres (*viz.* children's television, comedy, documentaries, fiction, and light entertainment). The main goal of this study is to demonstrate how subtitlers deal with the frequent use of Colloquial Belgian Dutch in the spoken source text on Flemish television. More specifically, it is investigated (i) to what extent Flemish subtitlers reproduce the spoken Belgian-Dutch colloquialisms in the subtitles or whether they even add colloquialisms to the subtitles (thereby enforcing the Belgian atmosphere), (ii) whether the subtitles contain more lexical than syntactic or morphological colloquialisms, (iii) whether the program genre influences these linguistic choices, and (iv) how the subtitlers themselves explain their linguistic choices. In order to achieve these goals, various lexical, syntactic, and morphological features were extracted from a parallel corpus, containing both the transcriptions of the spoken language in twenty television programs and the transcriptions of the corresponding subtitles. These data allow us to examine to what extent the original footage of the television program influences the subtitlers' linguistic choices. The outcome of this

study will subsequently give us more insight into the further spreading and acceptance of *tussentaal* Dutch in Flanders.

This paper is structured as follows. Section 2 gives a brief introduction into the contemporary language situation within Flanders, including its television landscape, and a concise overview of related linguistic research. In Section 3, the corpus materials and research method underlying this empirical case study are presented. The obtained results are discussed in Section 4, whereas the final section summarizes the major conclusions following from this investigation and gives some perspectives for future research.

2 Standard Dutch and the growing use of Colloquial Belgian Dutch ('tussentaal') in Flanders: a brief outline

2.1 Language situation in Flanders

The contemporary Flemish language situation is characterized by what Auer (2005/2011) names *diaglossia*. Unlike *diglossia*, which implies a language situation with on the one hand a codified, 'high' (standard) language variety and on the other hand one or more 'low' dialects (Ferguson, 1959; Fishman, 1967), *diaglossia* rather refers to a language continuum, consisting of various intermediate language varieties with features of both the standard language on the one hand and the dialects on the other. Nowadays in Flanders, a Belgian variety of Standard Dutch is widely accepted in the more formal registers and in written language (in newspapers, magazines, handbooks, official documents, on radio and television, etc.). This variety is also called *VRT-Dutch*, referring to the language variety that is used in informative radio and television programs on the Flemish public broadcaster. On the contrary, in less formal and informal contexts, however, language use in Flanders is characterized by typical lexical and grammatical features that are widely used, but not accepted as Belgian Standard Dutch (BSD) by the language authorities (e.g. Dutch Language Union, Hendrickx 1998). This non-standard informal variety is known as *tussentaal* (literally: 'in-between language'), and less often also termed Colloquial Belgian Dutch (CBD; e.g. Geeraerts & Van de Velde 2013; Ghyselen & Van Keymeulen 2016), a term that we will also use in this paper. Despite its name, however, it must be taken into account that *tussentaal* is not one coherent, homogeneous language variety, and that it is subjected to regional, social and even individual variation. Nevertheless, a couple of morphological, syntactic, and

phonological features have often been claimed to be 'core' elements of *tussentaal* (e.g. Rys & Taeldeman 2007; Taeldeman 2008). But even these features are proven not to be omnipresent in Flanders, and they do not have to be used per se in order to identify one's colloquial speech as *tussentaal*. These colloquial language varieties have been a prevalent topic in numerous heated discussions of Flemish linguists who shared the 'integrationist' vision (i.e. they advocated the adoption of the standard language of the Netherlands) and had been propagating the use of standard language for decades. This *tussentaal* was initially considered an inevitable transitional phase, an in-between language, used by speakers who are not yet capable to replace their dialect by the standard language, and which would ultimately disappear (e.g. Beheydt, 1993; Hendrickx, 1998). However, as soon as it became clear that this 'interlanguage' (Selinker, 1972; De Caluwe, 2005) was not a temporary phenomenon, but rather a widely used informal variety, it was described 'as ugly and deficient, as an intolerable, unnatural mix of two real, natural language varieties: the dialects on the one hand, and Standard Dutch on the other' (De Caluwe, 2005; see also Van Istendael, 1989; Van de Velde, 1996; De Schutter, 1998; Goossens, 2000; Taeldeman, 1992; De Caluwe, 2006). Although the *tussentaal* debate keeps alive until today, its focus has evolved from a reflective and evaluative point of view to a more empirical approach (Plevoets, 2013). From the beginning of the 21st century onwards, an increasing number of empirical studies started to investigate the use of CBD in various situational contexts, without passing judgement or even condemning this informal variety. For instance, De Caluwe (2009) studied the language used by young people (aged between 18 and 24 years) to outline their language behavior in informal conversations. Lybaert (2014a/2014b) set up an ethnographical study, in which her informants were asked to evaluate the spoken language in audio recordings in order to verify their attitude towards BSD and *tussentaal* (other attitudinal studies can be found in Impe & Speelman, 2007; Ghyselen, 2009 and Vancompennolle, 2012). The research of Delarue (2016) focuses on the position of *tussentaal* in educational contexts by analyzing the language use of school teachers. Other linguists attempted to define the form and content of *tussentaal* by compiling lists of its phonological and grammatical features (e.g. Rys & Taeldeman, 2007; Taeldeman, 2008).

In the abovementioned studies, it was repeatedly demonstrated that CBD has become increasingly common and accepted in various (mainly spoken) situational contexts, whereas BSD loses its status of the most appropriate variety. These results prove that the standard language ideology, which involves the dominant belief of 'one best language' (Swann et al.,

2004), has changed (Lybaert, 2015). Previous studies repeatedly attempted to attribute these changes in spoken Dutch to a process of destandardization or demotization (cf. Plevoets, 2008; Grondelaers & Van Hout, 2011; Grondelaers et al., 2011; Ghyselen, 2016; Delarue, 2016). Destandardization, is defined by Coupland & Kristiansen (2011, p. 28) as a context in which “the established standard language loses its position as the one and only ‘best language’”, which eventually leads to a greater tolerance towards language variation in all kinds of situations. Geeraerts & Speelman (2014) identify this as a process of dehomogenization. In a context of demotization, the standard language continues to be the ‘best’ variety, but with an extended interpretation of what this standard language exactly is, incorporating more variability (Mattheier, 1997). According to Grondelaers & Kristiansen (2013, p. 47), the standard language in this latter scenario continues to be the ‘best superiority language’ (e.g. in formal and written contexts), supplemented by a ‘best dynamism language’ (in informal spoken contexts). Unlike spoken Dutch, the traditional standard language ideology has stayed intact right up until today for the written discourse. These conditions have created in Flanders a linguistic dichotomy between a formal (mostly written) BSD variety on the one hand, and an informal (mostly spoken) CBD one on the other (Delarue, 2016, p. 25).

2.2 Language use on television: the Flemish public broadcaster VRT

Since its inception, the Flemish public broadcasting corporation VRT claims to be one of the most important distributors of BSD in Flanders. Its mission is to serve as an example of correct language use towards its audience. This norm-adherent language policy is written down in the VRT Language Charter (Hendrickx, 1998/2012), in which is stated that VRT aims to be the norm for the Belgian variety of Standard Dutch, stimulating its presenters to speak standard language. *Tussentaal* and dialect are only tolerated in radio and television programs in which it is functional, for instance, to retain the authenticity of the characters (e.g. in fiction and comedy). Newsreaders, reporters, journalists, voice-overs, and hosts in entertainment programs are expected to use standard language. Furthermore, special attention goes to the use of BSD in children’s television programs (Hendrickx, 1998/2012).

In the light of the increasing use of CBD in conversational and educational contexts (De Caluwe, 2009; Delarue, 2014), it does not come as a surprise that the language used in Flemish media has recently become a vexed question. For instance, Saman (2003) has demonstrated that the use of CBD in radio spots increased between 1991 and 2001. In addition, Van Gijssel

et al. (2004) have shown that language variation in radio and television advertising depends on format, medium, and target public. In their study, it was indicated that the use of *tussentaal* significantly increases in radio (vs. television) spots, in commercials intended for an adolescent target public and in dialogic minidramas. Lefevere (2011) and Prieels (2013) have found that, unlike the requirements of the VRT Language Charter (Hendrickx, 1998/2012), TV hosts frequently speak *tussentaal*. By analyzing the language used in the reality-show *Expeditie Robinson*, Zenner et al. (2009) have also shown that the participants adapt their language use to the conversational context, which implies that they are aware of the informal nature of CBD. In a diachronic study, Van Hoof (2013) has demonstrated that together with its educational role, the use of BSD in television fiction has decreased in a period of thirty years. Aforementioned studies have indicated that spoken language on Flemish television nowadays shows a lot of variation. Next to BSD, CBD is frequently used by TV hosts, and in commercials and entertainment programs. To ensure that the use of non-standard language varieties does not disturb the intelligibility of the television program, VRT provides intralingual subtitling. Intralingual subtitles are primarily intended for the deaf and hard-of-hearing audience, and are generally retrieved through teletext (i.e. closed subtitling) instead of appearing automatically on the screen (i.e. open subtitling). Only if the fragments are incomprehensible due to poor sound quality or murmuring, or when the speaker speaks dialect or a foreign language (e.g. English), the subtitles appear automatically on the screen (Hendrickx 2003; 2011). Given the wide distribution of CBD, not only in informal spoken contexts, but also in the spoken language of public institutions, the present study verifies whether this colloquial variety also infiltrates into the subtitles of the Flemish public broadcaster. Vandekerckhove (2007) and De Decker (2014), for instance, have demonstrated that CBD frequently occurs in chat conversations of Flemish teenagers. Consequently, it is particularly interesting to investigate which variety is used by subtitlers, since subtitles are written reproductions of spoken language with its typical colloquial features (Díaz-Cintas 2010).

For the subtitling practice on the Flemish public broadcaster, guidelines are written down in the VRT's guidebook for teletext subtitling (VRT, 2009). Next to instructions regarding the position and lay-out of the subtitles, this manual contains a number of guidelines concerning the use of standard and non-standard language. With regard to the lexicon, the style guide prescribes that *tussentaal* is reproduced in the subtitles as much as possible to retain the authenticity of both the program and the characters. Furthermore, if the TV host uses *tussentaal*, although he is considered to speak BSD, his

non-standard words are reproduced in the subtitles. *Tussentaal* morpho-syntax (e.g. the flexion of articles, adjectives, and pronouns) is always corrected. Recently, Prieels et al. (2015) and Prieels & De Sutter (2017) have indeed provided a first indication that Flemish subtitles contain colloquial words and constructions. In these studies, it was not only shown that the subtitlers use more non-standard language than translators of written text genres, but it was also revealed that subtitlers' linguistic choices are largely influenced by the source language, the speaker's type (voice over vs. actor/interviewee), and the program genre. First, the frequency of colloquial variants significantly increased in intralingual subtitles of Flemish speakers compared to interlingual subtitles of English speakers and intralingual subtitles of Netherlandic speakers. Secondly, if the subtitled speech came from an actor or interviewee (instead of a voice-over), the number of colloquial variants also increased significantly (Prieels et al., 2015). Finally, it was shown that the subtitles in informative documentaries and programs intended for children contained standard language to a large extent, whereas the number of colloquial lexemes and constructions increased significantly in the subtitles of fiction and comedy programs (Prieels & De Sutter, 2017). In earlier studies, also Remael et al. (2008) and De Ridder (2015) have demonstrated that intralingual Dutch subtitles contain significantly more non-standard lexicon than interlingual Dutch subtitles. These results suggest that Flemish subtitlers (consciously or unconsciously) reproduce the colloquial variants in the original footage of the television program in the subtitles. To confirm these assumptions, the influence of the spoken source text on the subtitlers' linguistic choices needs to be examined, which is the main purpose of this paper. In addition, this study attempts to detect whether the traditional belief that BSD is the only appropriate written language variety has changed. In other words, our research will provide further evidence as to whether the written discourse in Flanders is also characterized by a process of destandardization or demotization.

3 Methodology

3.1 Corpus materials

In order to examine to what extent the original footage of the television program influences the linguistic choices of the subtitlers, we built a parallel corpus, containing both the orthographic transcriptions of the spoken language in twenty Flemish television programs and the corresponding intralingual closed subtitles. The programs were transmitted by the

Table 1 Overview of the television programs in the corpus

Genre	TV program	Episode
Documentaries	Koppen	Operatie Vigilant Guard (01/06/2016)
		De zwarte lijst (08/06/2016)
	De vrije markt	18/06/2016
		25/06/2016
	Greece, the islands	Episode 1 (20/06/2016)
Episode 2 (21/06/2016)		
De Zevende Dag	05/06/2016	
	12/06/016	
Fiction	Thuis	Season 21, episode 4008 (01/06/2016)
		Season 21, episode 4013 (08/06/2016)
	De Ridder	Season 3, episode 12 (05/01/2016)
		Season 3, episode 13 (12/01/2016)
	T.	Season 1, episode 12 (06/12/2015)
		Season 1, episode 13 (13/12/2015)
Tom & Harry	Season 1, episode 9 (06/04/2015)	
	Season 1, episode 10 (13/04/2015)	
Children's television	Karrewiet	Episode 107 (30/05/2016)
		Episode 109 (01/06/2016)
	D5R	Season 3, episode 31 (01/06/2016)
		Season 3, episode 32 (02/06/2016)
	Klein gespuis	Season 1, episode 5 (30/05/2016)
		Season 1, episode 7 (1/06/2016)
Helden	Season 3, episode 2 (04/06/2016)	
	Season 3, episode 3 (11/06/2016)	
Comedy	Echt niet ok!	Episode 8 (09/02/2016)
		Episode 9 (16/02/2016)
	Nieuw Texas	Episode 6 (03/06/2015)
		Episode 7 (10/06/2015)
	Achter de feiten	Episode 9 (14/05/2014)
Episode 10 (21/05/2014)		
Light entertainment	Dagelijkse kost	Season 6, episode 1536 (30/05/2016)
		Season 6, episode 1537 (31/05/2016)
	Blokken	Episode 4694 (30/05/2016)
		Episode 4695 (31/05/2016)
	1000 Zonnen	(31/05/2016)
		(01/06/2016)
Over eten	(01/06/2016)	
	(08/06/2016)	

public broadcaster VRT (with the channels *Eén* and *Canvas*) between 2014 and 2016 and can be equally divided into five genre categories: children's

television, comedy, documentaries, fiction, and light entertainment. This division is largely based on Creeber's (2008) genre classification (cf. Prieels & De Sutter, 2017) for a detailed description of each genre). Each genre contains four different television programs of which each time two episodes of ten minutes were orthographically transcribed. In total, this corpus contains 103,808 words (speech: $n = 46,368$; subtitles: $n = 57,440$). The table below gives an overview of the different television programs that are included in our corpus.

3.2 Corpus data and method

In order to trace the linguistic choices of the subtitlers, various language features were extracted from our corpus. We are aware that the applied terminology is not completely accurate, as the BSD variants are in fact also used and accepted in the Netherlands, and have thus a General Standard Dutch status. However, as this study focuses on the Dutch language varieties used on Flemish television and VRT's language policy is oriented toward the use of a 'Belgian variety of Standard Dutch' (Hendrickx, 1998, p.1), we have chosen to adopt this terminology. Furthermore, General Standard Dutch features are automatically considered standard language in Belgium.

In a first step, the subtitles were manually scanned for colloquial features. If a colloquial item was found in the subtitles, the written transcription of the spoken source text was consulted to find the corresponding spoken variant. In addition, that same colloquial item was looked up in all speech transcriptions and the corresponding written alternatives of these spoken colloquialisms were checked to verify whether the subtitlers opt either for the reproduction of the colloquialisms or whether they convert the colloquialisms into BSD. Finally, all standard language variants were also searched for in the spoken source text to examine whether subtitlers also opt for a colloquial variant when a standard variant is used in the original speech. For example, the colloquial variant *appelsien* 'orange' was three times found in the subtitle corpus. In the transcriptions of the spoken source text, the corresponding spoken variants were in each case *appelsien*. Furthermore, an additional instance of *appelsien* was found in the transcriptions of the spoken corpus. In the corresponding subtitle, however, this spoken instance of *appelsien* was converted into the standard variant *sinaasappel*. Attestations of *sinaasappel* were not found in the transcriptions of the spoken source text. In other words, the selection process of the CBD feature *appelsien* resulted in 4 attestations in the spoken corpus, of which 3 instances were reproduced by the subtitlers and 1 instance was converted into the BSD alternative *sinaasappel*.

Table 2 List of the lexical features that were used in this study

Feature	Colloquial Belgian Dutch	Translation
1	accident	traffic accident
2	ajuin	onion
3	ambras	quarrel
4	appelsien	orange
5	boeleke	pet name for a baby
6	bol het af	get out
7	brol	trash
8	buizen	to flunk
9	camionette	delivery van
10	chance	luck
11	chapelure	breadcrumbs
12	chichi madam	chichi lady (negative connotation)
13	dagdagelijks	daily
14	efkes	just (temporal)
15	flik	cop
16	fretten	to scoff (food)
17	frigo	fridge
18	in het hol van Pluto	at the back of beyond
19	gelijk	like (comparison)
20	kostelijke affaire	an expensive deal
21	kozijn	cousin
22	kuisen	to clean
23	kuisvrouw	cleaning lady
24	madam	madam
25	nonkel	uncle
26	omwille van	because of
27	patat	patato
28	plezant	cheerful
29	saucisse	sausage
30	schoon	good-looking
31	seffens	soon
32	sjotten	play soccer
33	smissen	to make a mess of
34	stoefen	to brag
35	vijzen	to screw
36	weeral	again
37	eens	as soon as
38	zot	crazy
39	zever	twaddle
40	zwanzen	to joke

In total, 55 CBD features were extracted from the corpus. These linguistic items were divided into three types, viz. a set with lexical features (cf. Table 2),

Table 3 List of the morphological features that were used in this study

Feature	Colloquial Belgian Dutch	Translation
1	adjective (+e)	colloquial flexion of adjective
2	possesive pronoun (+e)	colloquial flexion of possessive pronoun
3	diminutive -ke	diminutive
4	ikke	flexion of the personal pronoun /
5	object <i>u</i>	object <i>you</i>

Table 4 List of the syntactic features that were used in this study

Feature	Colloquial Belgian Dutch	Translation
1	comparative + dan + object	comparative
2	durven + infinitive	<i>to dare</i> + infinitive
3	zijn ontslag geven	<i>to resign</i>
4	niet moeten	<i>not have to</i>
5	preposition + preposition	preposition + preposition
6	zijn	possesive <i>zijn</i>
7	onze/ons + proper name	<i>our</i> + proper name/generic name
8	de + proper name	<i>de</i> + proper name
9	zet je erbij	<i>have a seat</i>
10	aux + part + inf	position of the participle in the verbal end group

a set with morphological features (cf. Table 3), and a set with syntactic features (cf. Table 4). Appendix 1 contains a representative selection of corpus examples of each of the linguistic features.

The selection of these linguistic features is based on a number of normative sources. To verify whether a variant was labelled CBD, we consulted Van Dale (2015), *Taaladvies*² and lists of tussentaal features compiled by De Caluwe (2006) and language advisor Hendrickx (2001). Features were only selected if all three sources were unanimous in characterizing them as standard or non-standard language. It should be noted, however, that the status of a couple of these features is nowadays disputed. The possessive construction with *zijn* (e.g. *Pa zijn camionette* ‘Fathers’ delivery van’), for instance, is accepted by some sources in informal written language. Nevertheless, we have chosen to add these items to our lists with colloquial features, as they are still not considered BSD.

One of the consequences of the methodology that we applied, is that we only selected the colloquial features that were reproduced in the subtitles and that other typical ingredients of *tussentaal* (e.g. the personal pronoun *ge/gij* and reduplication of the subject) were not included. Although these features did not appear in the subtitles, it could be interesting to know how

often subtitlers do convert these typical CBD elements into BSD. To solve this shortcoming, we consulted lists of *tussentaal* features, compiled by several Dutch linguists (De Caluwe, 2006; Everaert, 1998; Geeraerts et al., 2000; Hendrickx, 2001; Lebbe, 1996; Taeldeman, 2008; Van Gijssel, 2008) in the past few decades. We selected nine features that are considered as typical ingredients of CBD and added them to our dataset in order to investigate how often subtitlers convert these elements into BSD. An overview of these CBD features is given in Table 5.

Table 5 List of additional *tussentaal* features that were used in this study

Feature	Colloquial Belgian Dutch	Translation
Morphological features		
1	verbuiging lidwoord: <i>den, ne(n)</i>	flexion article
2	verbuiging aanwijzend voornaamwoord	flexion demonstrative pronoun
3	<i>ge/gij</i>	personal pronoun <i>you</i>
4	1 st e persoon enkelvoud + <i>n</i>	colloquial conjugation 1st p. singular
5	2 ^e persoon enkelvoud	colloquial conjugation 2nd p. singular
Syntactic features		
6	<i>van/voor</i> + beknopte bijzin	reduced clause
7	dubbele negatie	double negative
8	redundant <i>dat</i>	redundant <i>that</i>
9	subjectsreduPLICATIE	reduplication of the subject

The data extraction and validation resulted in a final dataset of 1756 relevant Colloquial Belgian Dutch attestations (spoken: $n = 1616$; subtitles: $n = 140$) and 1490 relevant Belgian Standard Dutch alternatives (spoken: $n = 0$; subtitles: $n = 1490$).

3.3 Quantitative analysis

To verify whether the original, colloquial footage of the television program influences the linguistic choices of the subtitlers, we calculated the proportion between CBD and BSD in the subtitle corpus in relation to the use of CBD in the spoken corpus. First, we counted the CBD variants in the spoken corpus. Next, both the CBD variants and the BSD variants in the corresponding subtitles were counted. These absolute scores of CBD variants and BSD variants in the subtitle corpus were then divided by the total number of CBD variants in the spoken corpus. This resulted in a relative frequency score which indicates how often subtitlers reproduce the CBD variants of the spoken source text in the corresponding subtitles or how often they translate those CBD variants into BSD. This

calculation was made for the three feature sets separately with the intention of comparing the use of colloquial and standard lexical, morphological, and syntactic variants by subtitlers. Additionally, similar CBD and BSD calculations were made for each program genre separately in order to verify whether the program genre influences the subtitles' linguistic behavior. Because of the small size of our corpus materials, the frequency scores of the individual language variants will be left aside in the following section.

3.4 Qualitative analysis

In order to enhance the value of this research, we performed a qualitative analysis of the obtained results in the quantitative study, supplemented with semi-structured interviews, evaluation reports, and observational data that were collected at VRT's subtitling department. As Ghyselen et al. (2016) mention, it is useful to supply our data on production with some data on perception. Furthermore, we want to get a better insight into the context in which subtitlers are working. Looking at the process of linguistic data is a successful method that has been introduced in the studies on news production of Jacobs et al. (2011). The ethnographic approach in the present study will provide us with more contextual information, foregrounding some of the practical as well as political concerns subtitlers have to deal with (Sleurs & Jacobs, 2005). The underlying idea of this qualitative study is to get more insight into the subtitlers' attitude towards VRT's language policy on the one hand and the language reality in Flanders on the other hand, which is crucial to understand their linguistic choices in the subtitles.

In a first step, we interviewed the head of Translation and Subtitling at VRT about the language policy of the public broadcaster in relation to the current linguistic situation and the increasing use of *tussentaal* on Flemish television. This meeting gave us a detailed insight into the guidelines VRT's subtitlers are particularly bounded by. Next, we organized an interview with two subtitlers: the first interviewee (subtitler 1) is currently working at the editorial board of the T888-department (intra-lingual closed subtitling); the second interviewee (subtitler 2) is currently working at the editorial board of translations and interlingual subtitling.³ Both subtitlers are aged between 25 and 30 and they are both employed at VRT since 2012, after having completed their Master's in Translation and Interpreting. The interviews took place in the office of the subtitlers and consisted of (i) a standardized questionnaire to document information

about the participants (e.g. age, education, geographical background) and (ii) a series of more or less structured questions concerning the subtitlers' profession, their vision on language ideology and their opinion towards the use of CBD in some corpus subtitles of this study. Each interview lasted about sixty minutes, so for the sake of the interviewer, the conversations were recorded. After having interviewed the subtitlers, our data were enriched by means of participant observation (Brannan & Oultram, 2012; Duranti, 1997) to get more detailed insight into the subtitling process. We observed subtitler 1 while she was subtitling an episode of the fiction series *Thuis* (season 22, episode 4185) and we asked her to produce a so-called 'continuous concurrent protocol' (Ericsson & Simon, 1993), during which she explicated her choice for a specific word or construction in the course of the writing process. We made extensive field notes on what the subtitler did and said in the belief that listening is the best strategy for learning (Myers, 1986).

4 Results and discussion

4.1 Quantitative analysis

In this section, the results of our quantitative analysis are presented and discussed. First, we will examine to what extent Flemish subtitlers transfer the spoken Belgian-Dutch colloquialisms to the subtitles and whether there are differences between lexical, morphological and syntactic variants, thus answering our first and second research question. Next, we will focus on the influence of the program genre on the subtitlers' linguistic choices in order to answer our third research question.

4.1.1 General observations

When initially compiling our dataset, we only selected the CBD features that appeared in the subtitles. As a consequence, some typical *tussentaal* elements were not included in our dataset. To solve this shortcoming, we consulted existing lists of *tussentaal* ingredients and added a set of nine commonly used CBD features to our data. Even though these elements did not occur in the subtitles, it is interesting to examine how frequently subtitlers convert them into standard language. Table 6 shows how many times these CBD items were used in the spoken source text and, consequently, how often the subtitlers converted these colloquialisms into BSD.

Table 6 Total number of attestations of typical CBD features in the spoken source text

Feature	Colloquial Belgian Dutch	Attestations in spoken source text
Morphological features		
1	flexion article: <i>den, ne(n)</i>	227
2	flexion demonstrative pronoun: e.g. <i>dieje(n), diene(n), dezen</i>	63
3	personal pronoun <i>ge/gij</i>	461
4	colloquial conjugation 1st person singular (+ <i>n</i>)	23
5	colloquial conjugation 2nd person singular	86
Syntactic features		
6	<i>van/voor</i> + reduced clause	6
7	double negative	7
8	redundant <i>that</i>	79
9	reduplication of the subject	117

In Table 6, three features immediately attract attention because of their high number of occurrence in the spoken corpus: the personal pronoun *ge/gij* ($n = 461$), flexion of the article ($n = 227$), and reduplication of the subject ($n = 117$). It does not surprise, on the one hand, that subtitlers do not reproduce these CBD features in the subtitles, since VRT’s guidelines prescribe that morphosyntactic *tussentaal* elements should be converted into standard language. On the other hand, it will be shown later on in the analyses that other morphological and syntactic colloquialisms (cf. Table 3 and 4) do occur in the subtitles, although subtitlers are expected to translate them into BSD. Furthermore, the total number of attestations of each of the CBD features in Table 6 is significantly higher than the total number of attestations of the reproduced colloquialisms in Table 3 and 4 (see Appendix 2 for the total number of CBD and BSD attestations per dataset in the subtitles). Nevertheless, all features in Table 6 were converted into BSD in the corresponding subtitles, regardless of their frequent occurrence in the spoken source text. Based on the interviews with the subtitlers (cf. Section 4.2.), it turned out that, for instance, the use of the personal pronoun *ge/gij* also requires a CBD variant of the following verb (e.g. *ge zijt* and not *ge bent* ‘you are’). Although the personal pronoun *ge/gij* is a very commonly used colloquial feature, VRT does absolutely not approve the reproduction of the CBD variant of the verbal form (*zijt*). As a consequence, both the personal pronoun *ge/gij* and the verbal form *zijt* are converted into BSD, viz. *je/jij bent* ‘you are’. In Section 4.2., it will be explored in depth why subtitlers opt either for the reproduction or for the conversion of the CBD features.

As we already mentioned, CBD features occasionally do appear in subtitles on Flemish television. a first look into our corpus data demonstrated

that subtitlers use these colloquialisms only if the spoken source text contains such a colloquial feature. In other words, subtitlers will never add a CBD feature to the subtitles if it does not appear in the original speech. For that reason, we will only incorporate the spoken Belgian-Dutch colloquialisms in the following analyses in order to find out to what extent they are transferred to the subtitles. Additionally, we will leave out the data of Table 6, because these *tussentaal* items did never occur in the subtitles. From this point on, we will merely focus on the features in Tables 2, 3, and 4 of which we know that subtitlers did use them. By doing so, we will not only be able to verify how often these spoken CBD variants are reproduced in the subtitles, but we will also reveal whether there are differences between lexical, morphological, and syntactic features.

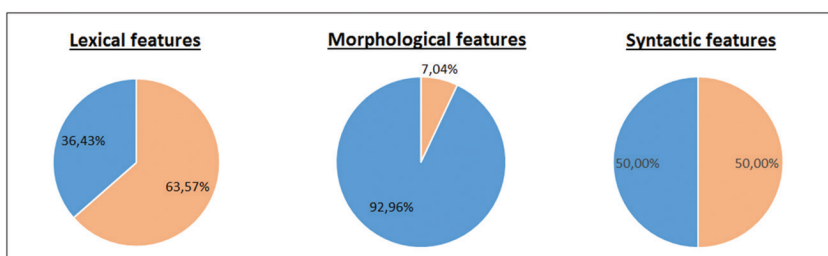


Figure 1 Use of Colloquial Belgian Dutch (light grey) and Belgian Standard Dutch (dark grey) in the subtitles

Figure 1 shows the proportion between CBD and BSD variants that were used by the subtitlers for each feature set. The relative frequency of CBD is marked in light grey, whereas the relative frequency of BSD is marked in dark grey. The diagrams reveal two main findings. First, it can be observed that subtitlers make use of CBD variants in intralingual subtitles on Flemish television. These findings confirm the results of several recent studies into linguistic variation in Flemish subtitling (e.g. De Ridder, 2015; Prieels et al., 2015; Prieels & De Sutter, in press). Secondly, the diagrams show that subtitlers especially reproduce CBD lexemes in the subtitles, whereas syntactic and, in particular, morphological colloquialisms are more often translated into standard language. For the lexical features, 89 of the 140 spoken Belgian-Dutch colloquialisms (63,57%) were reproduced in the subtitles, whereas 51 spoken CBD lexemes (36,43%) were converted into a BSD alternative. This outcome is not quite surprising, since the Flemish public broadcaster promotes the reproduction of colloquial lexicon to retain the authenticity of the television program and the characters. Nevertheless, the diagram shows that subtitlers in a significant number

of cases (36.43%) do translate the CBD lexemes into a BSD alternative. An in-depth analysis of the dataset reveals that these results are largely influenced by the disproportion of three individual features. The CBD lexemes *efkes* ('just'), *gelijk* ('like'), and *seffens* ('soon') are merely once reproduced in the subtitles. Apart from that, these features are converted into standard language (respectively 32, 7, and 4 times), which could explain the strong presence of BSD lexemes in the diagram. Other colloquial lexemes that were always reproduced in the subtitles (e.g. *ajuin* 'onion' and *brol* 'rubbish') merely appeared once in the corpus, which intensifies even more this disproportion between these two 'groups' of lexical colloquialisms. For the reproduced CBD lexemes (i.e. the light grey part of the diagram), on the contrary, there is no such disproportion between the individual variants. To solve the disproportion that is caused by *seffens*, *gelijk*, and *efkes*, we calculated the normalized frequencies for all lexical variants. As such, each lexeme was assigned an equal weight, regardless of their frequency of appearance in the subtitles. The resulting diagram shows that subtitlers generally reproduce the CBD lexemes in the subtitles (89.96%), whereas only 10.04% of the lexical colloquialisms are converted into a BSD alternative (the diagrams with the normalized frequencies can be consulted in Appendix 3). In other words, the calculation of the normalized frequencies of the lexemes confirms that the results in Figure 1 are largely influenced by the disproportion of *seffens*, *gelijk*, and *efkes*.

Given the peculiar attention for these three lexemes, the question arises why subtitlers avoid *seffens*, *gelijk*, and *efkes* in particular. To get more insight into the reason behind their linguistic choices, we set up a continuum with the variants that are reproduced in the subtitles at the left pole and the variants that are converted into BSD at the right pole. Next, the CBD lexemes with at least five attestations in our corpus were distributed along the continuum, with their position depending on how often they were reproduced or converted into BSD in the subtitles.

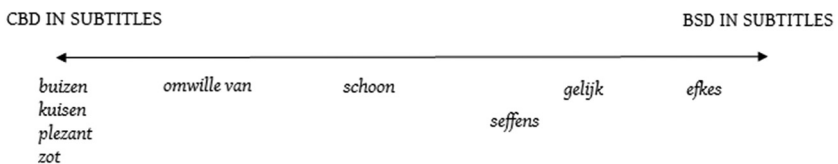


Figure 3 Continuum that visualizes the reproduction/conversion of the lexical features in this study

Seffens, *gelijk*, and *efkes* are situated at the right side of the continuum, since these lexemes are generally converted into BSD in the subtitles. *Buisen* ('to

flunk'), *kuisen* ('to clean'), *plezant* ('cheerful'), and *zot* ('crazy'), on the contrary, are clustered at the left pole of the continuum, as these CBD lexemes were reproduced in the subtitles in 100% of the cases. *Omwille van* ('because of') and *schoon* ('good-looking') are distributed along the continuum. Although these lexemes are similarly labelled as CBD, subtitlers seem to cope differently with them. Probably, *seffens*, *gelijk*, and *efkes* are more salient than the other lexemes, which means that they are easily detected as CBD by language users. This could explain why subtitlers generally replace these lexemes by their BSD alternative. The salient nature of *efkes* can be attributed to the inclusion of the *-ke* diminutive. In her study, Lybaert (2014) has demonstrated that the *-ke* diminutive is a salient morphological feature, since language users generally refer to this element as a typical feature of *tussentaal*. Based on the interviews in Section 4.2, we will discuss this salience effect, together with other reasons for the frequent conversion of *seffens*, *gelijk*, and *efkes*.

For the morphological features, Figure 1 shows that the subtitlers copied only 28 of the 358 spoken CBD morphemes (7.82%) to the subtitles, whereas 330 spoken CBD features (92.18%) were replaced by a BSD alternative. Just like the outcome of the lexical features, the results for the morphological colloquialisms match our expectations. In its style guide, VRT prescribes that morphosyntactic *tussentaal* items should always be corrected, so it does not come as a surprise that subtitlers generally convert these CBD morphemes into BSD variants. However, in 7.82% of the cases, the colloquial morphemes are reproduced by the subtitlers. We also calculated the normalized frequencies for the morphological set to eliminate a potential disproportion between the individual variants. The resulting diagram shows, however, that this calculation does not have a great influence on the results (cf. Appendix 3 for the diagrams with the normalized frequencies). The continuum below visualizes the occurrence of the individual CBD morphemes in the subtitles.



Figure 4 Continuum that visualizes the reproduction/conversion of the morphological features in this study

This profound look into our data indicates that out of the 28 reproduced CBD morphemes especially the diminutive *-ke* ($n = 14$) and the informal object *u* ($n = 9$) are copied to the subtitles. These frequency differences between the individual morphological items are not atypical, since some morphemes are considered to be more salient than other (Van Bree, 2000). However, it is remarkable that especially the diminutive *-ke* and the informal object *u* are reproduced by the subtitlers, whereas Lybaert's (2014) study has demonstrated that these linguistic items in particular are two salient language features. In other words, language users easily detect these items as CBD variants. Although the mayor part of the CBD morphemes of the spoken source text are converted into BSD, the subtitlers 14 times opt for the reproduction of the diminutive *-ke* and 9 times for the reproduction of the informal object *u*. These numbers are remarkably high, albeit it can be assumed that subtitlers are aware that these morphemes are *tussentaal*. During the interviews, these results were presented to the subtitlers to clarify this remarkable outcome. In Section 4.2. we will discuss their comments in detail.

Finally, Figure 1 shows that 26 of the 52 spoken CBD constructions (50.00%) were reproduced in the subtitles, whereas 26 spoken CBD constructions (50.00%) were converted into BSD constructions. Furthermore, when we calculate the normalized frequencies for each variant to assign them an equal weight, it turns out that the proportion of CBD constructions in the subtitles even increases to 60.74% (cf. Appendix 3). Contrary to the lexical and morphological features, the frequent occurrence of syntactic colloquialisms in the subtitles does not match our expectations, since VRT's style guide prescribes that CBD syntactic constructions should always be corrected. An explanation for this remarkable observation can be found in the study of Lybaert (2014), in which it was shown that the syntactic domain is less salient than the lexical domain. According to Van Bree (2000), this can be attributed to the abstract nature of the syntactic elements: unlike the lexicon, this domain is characterized by abstract rules, which makes it more automatized or less concrete, so language users unconsciously use these features. When we put the syntactic features with at least five attestations along the continuum below, it can be seen that subtitlers cope differently with the individual variants

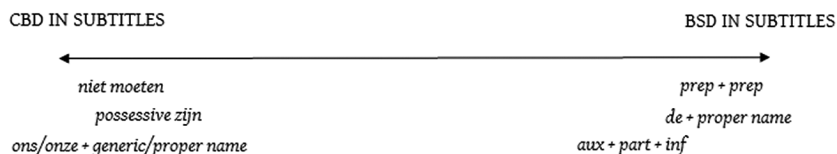


Figure 5 Continuum that visualizes the reproduction/conversion of the syntactic features in this study

The continuum demonstrates that subtitlers frequently transfer three colloquial features in particular: *niet moeten* ('do not have to'), the possessive *zijn*, and *ons/onze* ('our') + proper name. Constructions with two (uninflected) prepositions, *de* + proper name, and *aux* + *part* + *inf* are generally translated into BSD. The frequent reproduction of *ons/onze* + proper name can be explained by the subtitling guidelines of VRT. In its style guide, the public broadcaster prescribes that the flexion of pronouns is not allowed in the subtitles, with the exception of the construction *ons/onze* + proper name. Other syntactic colloquialisms need to be converted into standard language. Assuming that, as language professionals, subtitlers are aware that *niet moeten* and possessive *zijn* are CBD features, there have to be other underlying arguments which explain why subtitlers opt for the reproduction of these colloquialisms in particular. In Section 4.2. we will discuss the subtitlers' comments on this issue.

Although this first analysis revealed that lexical colloquialisms are more often reproduced in intralingual subtitles on Belgian television than morphological and syntactic colloquialisms, it merely gives an idea of the subtitlers' linguistic choices in general, without taking into consideration program genre differences. Furthermore, our results showed that subtitlers in a remarkable number of cases do opt for the reproduction of CBD morphemes (7.82%) and constructions (50.00%), although *tussentaal* morphosyntax should always be corrected according to the style guide. As a consequence, the question arises by which factors the subtitlers are driven to reproduce these morphological and syntactic colloquialisms. In a previous study (Prieels et al., 2015), we have studied the effect of the program genre on the subtitlers' linguistic choices, which showed that subtitles in informative programs contain standard language to a large extent, whereas the number of CBD lexemes and constructions increased significantly in the subtitles of entertainment programs. Therefore, we will focus in the next paragraph on the influence of the contextual parameter program genre with regard to the use of CBD or BSD in the subtitles. By

doing so, we will be able to answer the question whether the language choices made in subtitles produced for Flemish television differ according to the program genre.

4.1.2 *The influence of (program) genre on the subtitlers' linguistic choices*

In a recent study (Prieels & De Sutter, 2017), we have demonstrated that the intralingual subtitles of Belgian Dutch television programs show a lot of genre variation, as certain genres (*fiction* and *comedy*) contain more CBD in the subtitles than other genres (*documentaries* and *children's television*). In this section, it will be indicated in which genres the subtitlers reproduce the Belgian-Dutch colloquialisms of the spoken source text in the corresponding subtitles and whether their choices differ depending on lexical, morphological, and syntactic variants.

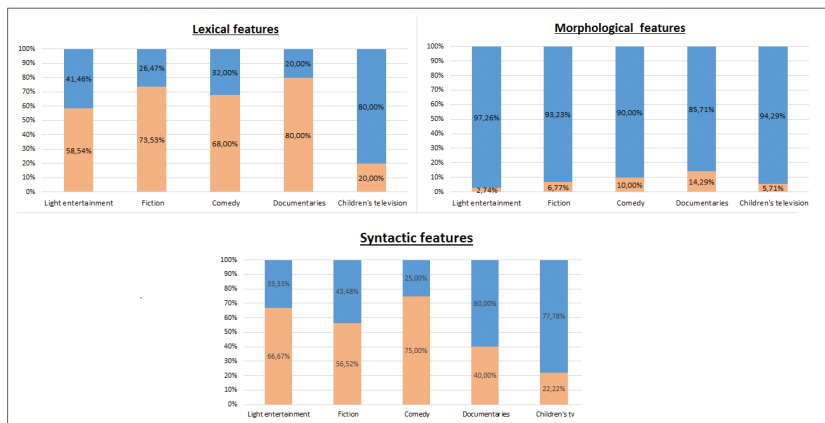


Figure 6 Use of Colloquial Belgian Dutch (light grey) and Belgian Standard Dutch (dark grey) in the subtitles of five program genres

In Figure 6, the proportion between CBD and BSD variants in the intralingual subtitles of five program genres is presented for each feature set. Once again, it can be observed that subtitlers often transfer the CBD lexemes and constructions in the spoken source text to the subtitles, whereas they generally translate the colloquial morphological elements to BSD (cf. supra). Furthermore, the diagrams reveal genre differences with regard to the use of BSD and CBD, especially with regard to the lexical and syntactic features. For the lexical variants, it can be seen that the spoken CBD lexemes are more often reproduced than converted into BSD in the subtitles of *light entertainment* (58.54%), *fiction* (73.53%), and *comedy* (68.00%), whereas in

the subtitles of *children's television*, 80.00% of the spoken CBD lexemes are replaced by a BSD alternative. In the subtitles of *documentaries*, the number of CBD lexemes is also remarkably higher (80.00%) than the number of BSD lexemes (20.00%). In this genre, however, the small number of attestations ($n = 5$) could give a distorted picture. Although the relative number of CBD lexemes is the highest in the subtitles of *documentaries*, the absolute number of CBD lexemes is more than six times higher in the subtitles of *light entertainment*, *fiction*, and *comedy* (cf. Appendix 2 for the total number of CBD and BSD attestations per dataset in each genre). In the diagram with the morphological features, the influence of the program genre is less outspoken. The morphological colloquialisms are generally converted into a BSD alternative and the higher amount of CBD in the subtitles of *documentaries* (14.29%) can be attributed to the small number of attestations ($n = 15$), which prevents us of generalizing these results. Even though the relative number of CBD morphemes is the highest in the subtitles of *documentaries*, their total number of CBD morphemes is even lower than in the subtitles of *fiction* and *comedy*. The diagram with the syntactic features shows that the spoken CBD variants are more often reproduced than converted into BSD in the subtitles of *comedy* (75.00%), *light entertainment* (66.67%), and *fiction* (56.52%), whereas in the subtitles of *children's television* and *documentaries*, the spoken CBD constructions are more often replaced by a BSD alternative (resp. 77.78% and 60.00%).

The higher number of CBD lexemes and syntactic constructions in subtitles of *comedy*, and to a certain extent in those of *fiction* and *light entertainment*, can be explained by the general aim of those programs. The main objective of laughing and entertaining programs is to amuse the audience by creating an informal, spontaneous atmosphere which has a greater chance of showing spontaneous, colloquial utterances (e.g. McIlveny et al., 1993; Remael, 2003; Rutter, 1997). Furthermore, the public broadcaster tolerates the use of CBD in favor of the authenticity of entertainment programs. As a result, the subtitlers seem to reproduce the colloquial features in the subtitles to retain this informal, spontaneous nature of the TV program and to avoid that 'the characters speak like a printed page' (Rosa, 2001, p. 216). Conversely, the educational role of children's television could explain why subtitlers translate the CBD variants into BSD. Children's programs aim to perform an exemplary role (Nikken & Friebel, 1990), also on the level of language use, which results in the recurring use of BSD, both in the spoken source text and in the subtitles. Furthermore, in its Language Charter, the public broadcaster VRT emphasizes the importance of standard language in children's programming (Hendrickx, 1998/ 2012).

4.2 Qualitative analysis

The quantitative analyses in the previous paragraph yielded some interesting results with regard to the use of Belgian-Dutch colloquialisms in subtitles on Flemish television. Not only have we learnt that subtitlers more often reproduce CBD lexemes than CBD morphological and syntactic features, but it has also become clear that the program genre influences the subtitlers' linguistic choices. Furthermore, we have seen that subtitlers do neither consequently reproduce the colloquial lexemes, nor do they convert every morphological or syntactic colloquialism into BSD. The diminutive *-ke* and the informal object *u*, for instance, are frequently copied to the subtitles, whereas the lexemes *efkes*, *seffens*, and *gelijk* are almost every time replaced by a BSD alternative. As a result, the obtained findings require some more clarification. Therefore, we have interviewed the head of the subtitling department and two subtitlers at VRT. Subtitler 1 currently works at the editorial board of the T888-department (intralingual closed subtitling) and subtitler 2 currently works at the editorial board of translations and interlingual subtitling. Initially, both departments worked separately, but nowadays the distinction between open and closed subtitling is less definite, since open subtitling largely adopts the procedures of the closed subtitling department. In addition, we observed subtitler 1 while she was subtitling an episode of the fiction series *Thuis* (season 22, episode 4185). At VRT's subtitling department, each subtitler manually subtitles a couple of TV programs a day. During the subtitling process, subtitler 1 regularly consults the Van Dale dictionary to verify the normative status of a word. *Taaladvies* is used for the verification of grammatical constructions. When the subtitler has finished, the subtitles are sent to a colleague for the final editing. In the following sections, we will first focus on the subtitlers' vision concerning VRT's subtitling guidelines. Next, we will discuss their comments on some corpus examples, thus answering our fourth research question. In the discussion, quotes of the interviewees are translated by the authors of this paper and put between quotation marks.

4.2.1 *The subtitlers' perceptions of the subtitling policy at VRT*

Based on the interviews, it became clear that the subtitlers attach great value to VRT's subtitling instructions. With regard to the lexicon, Van Dale dictionary is their main authoritative source. During the subtitling process, it was remarkable how often subtitler 1 consulted Van Dale dictionary to verify which label was ascribed to certain words: 'If Van Dale dictionary labels a word as "colloquial" or "informal", this lexeme must be reproduced in the subtitles'. Only in case of spatial or temporal restrictions, a subtitler will

deviate from this rule (e.g. *ongeval* ['accident'] counts less characters than *accident* [accident]). Even interjections like *allee*, *amai*, *voilà*, *hé*, and *huh* are reproduced in the subtitles, because Van Dale dictionary labels them as CBD. According to the head of the subtitling department, VRT's tolerant attitude toward the use of colloquial lexicon is mainly based on the needs and requirements of the deaf and hard-of-hearing audience. They want the subtitles to stay as close as possible to the spoken source text, not only to match the mouth image, but also to expand their knowledge of the Dutch language. Nevertheless, subtitler 1 admits that although the public broadcaster promotes the reproduction of colloquial lexicon, she would only use, for example, *verschieten* ('to be frightened') in the subtitles of a fiction series, whereas she would replace this colloquial lexeme by its BSD alternative *schrikken* in a documentary. In fiction series, the use of colloquial varieties serves a useful purpose, which is to create an authentic atmosphere. These colloquial varieties do not have such a function in other program genres, so the subtitlers 'clean them up'.

With regard to morphosyntactic constructions, the subtitlers defend VRT's strict guidelines: 'Colloquial grammatical constructions are absolutely not allowed in the subtitles'. Even if the BSD construction is too long to fit into the frame and the CBD variant counts less characters, subtitlers will always rephrase the sentence until they have formulated a grammatically correct subtitle. One of their main concerns is the audience. According to the subtitlers, it would undoubtedly raise an avalanche of complaints if the subtitles contained grammatical mistakes. Subtitler 2 admits, however, that in the last years, the instructions concerning grammatical correctness have changed. *Taaladvies*, for example, has become more tolerant toward certain grammatical issues. Linguistic items that were disapproved before (e.g. *dit keer* 'this time') are nowadays considered as standard language. Although subtitler 2 does not always agree with this, 'as a subtitler, he has to put his feelings aside and follow the instructions'. This proves once again that the subtitlers attach great importance to these subtitling guidelines.

At the end of the interview, both subtitlers indicate that they fully support the linguistic guidelines as they are formulated in VRT's style guide. Although these instructions must be followed in theory, the head of the subtitling department emphasizes that, in actual practice, the subtitlers merely 'aim to follow the official subtitling guidelines'. Due to pressure of time, for instance in live subtitling and last-minute translations, 'errors can occasionally be found in the subtitles'. In order to eliminate language mistakes, the majority of the subtitles are submitted to a final editing process before they appear on the screen. To alert the subtitlers to their mistakes,

the subtitling department has recently introduced a feedback system. After the revision by the editorial board, a document with corrections and comments is sent to all editors, including the responsible subtitler. By doing so, VRT hopes to inform its subtitlers about frequently made errors in order to make them avoid similar mistakes in the future. The system turned out to be successful, since the subtitlers think it is very helpful, especially to supervise beginner colleagues.

4.2.2 *The subtitlers' perceptions of the subtitling reality at VRT: some corpus examples*

The results in Section 4.1.1. have shown that lexical colloquialisms are more frequently used in the subtitles of Flemish-spoken television programs than syntactic and, particularly, morphological colloquialisms. These findings are illustrated with some corpus examples below. Each example consists of the original spoken text and the TV program between brackets (first line), the corresponding subtitle (second line) and the translation in English (third line). The CBD features that were incorporated in our study are marked in bold; other colloquialisms are marked in italics.

1. Een **kuisvrouw** kost geld, alles kost geld tegenwoordig. (1000 Zonnen)
Een **kuisvrouw** kost geld, alles kost geld tegenwoordig.
A **cleaning lady** costs money, everything cost money nowadays.
2. *Hoeda?* Is z' **ambras** komen maken? (Thuis)
Hoe? Is ze **ambras** komen maken?
How? Did she come here to make a **quarrel**?
3. Onze papa deed dat ook met *den* **ajuin**. (Dagelijke kost)
Onze papa deed dat ook met **ajuin**.
Our father also did that with **onion**.
4. Is er al cava? Ah, ja. *Wilde* mij *is* een **glaske** geven? (Echt niet ok!)
Is er al cava? Ah, ja. Wil je mij eens een **glaasje** geven?
Do you have cava? Could you give me a **glass**?
5. Om **de Luc** e plezier te doen, *da* 's alles. (Thuis)
Om **Luc** een plezier te doen. Dat is alles.
To do **Luc** a favor. That's all.

In example 1, 2, and 3 the spoken CBD lexemes *kuisvrouw* ('cleaning lady'), *ambras* ('quarrel'), and *ajuin* ('onion') are reproduced in the subtitles, whereas the CBD *-ke* morpheme in example 4 and the CBD construction *de Luc* in example 5 are converted into a BSD alternative. When we present

these corpus examples to the subtitlers, they declare that the five examples are completely in accordance with the subtitling guidelines. First, the colloquial *-ke* diminutive and the construction *de Luc* are both corrected, since these variants are not accepted by the official language advices of the Dutch Language Union, on which VRT's subtitling guidelines are based. Although *glaske* is very commonly used in everyday speech, it will never appear in written language, so according to the interviewees, 'it would be odd to read this *-ke* diminutive in the subtitles'. As a consequence, subtitlers will always use the BSD variant *glaasje*. Secondly, words like *kuisvrouw*, *ambras*, and *ajuin* which are labelled as CBD in Van Dale dictionary are allowed by VRT's style guide. According to the subtitlers, converting these lexemes into standard language could even cause a disturbing effect, since the subtitles would then deviate too much from the spoken source text. Furthermore, not only are *kuisvrouw*, *ajuin*, and *ambras* 'commonly used in spoken conversations, *kuisvrouw* as well as *ajuin* do also frequently occur in written language (e.g. in recipes or advertisements)'. According to subtitler 1, these lexemes are sufficiently intelligible to the audience and 'they do not "feel" Flemish', which supports them to be used in the subtitles.

In the aforementioned corpus examples, there are some CBD features that were not transferred from the spoken source text to the subtitles. The deletion of *-t* (*da[t]* 'that'), the end vowel (*z[e]* 'she'), and the first vowel (*[i]s* 'is') as well as the colloquial form of the personal pronoun (*wil*)*de* ('could you') are translated into BSD. Subtitler 1 explains that the deletion of vowels (*[i]s* 'is' and *[ee]ns* 'once') only occasionally occurs in open intralingual subtitling. This can usually be attributed to spatial and temporal restrictions, since subtitlers are bounded by the so-called six-second rule, which involves that television viewers are able to read two-lined subtitles with a maximum of 70 to 74 characters in a time span of six seconds (Díaz Cintas & Remael, 2014). In closed intralingual subtitling, on the other hand, this vowel reduction is not usual, because it could cause confusion to the deaf and hard-of-hearing audience. The *-t* deletion in *da[t]* *[i]s* ('that is') is 'too dialectal', so subtitlers will never reproduce this colloquial feature in the subtitles. The same goes for the colloquial form of the personal pronoun *je/jij*. In CBD, *je/jij* can be produced as *ge/gij* or as an enclitic *-de* (e.g. *wilde*). According to subtitler 2, these colloquial variants are never used in the subtitles, because they will often cause grammatical issues. The personal pronoun *ge/gij*, for instance, requires a colloquial conjugation of the verb (e.g. *ge waart* and not *ge was* 'you were'). The reproduction of this CBD variant of the verbal form (*waart*) is absolutely not approved by VRT. Furthermore, the personal pronoun *ge/gij* rarely appears in written language (with the exception of

chat and text messages), even though it is very commonly used in spoken conversations. As a consequence, subtitlers will always convert the personal pronoun *ge/gij* and the verbal form *waart* into standard language, viz. *je/jij was* ('you were').

With regard to the lexical features, the quantitative analyses have also indicated that the colloquial lexemes *efkes* ('just'), *gelijk* ('like'), and *seffens* ('soon') were merely once reproduced in the subtitles, whereas in all other cases they were converted into a BSD alternative (cf. some corpus examples below).

6. **Efkes** serieus blijven, hè. (Achter de feiten)
Even serieus blijven, hè.
Just be serious.
7. Ik heb twee keer gehuild **gelijk** een klein kind. (Echt niet ok!)
Ik heb twee keer gehuild **als** een klein kind.
I cried **like** a baby twice.
8. En dan kunnen we **seffens** als die in *den* oven steken. (Dagelijkse kost)
En dan kunnen we **straks**, als ze in de oven steken...
And **later**, when they are in the oven, we can ...

On the one hand, the subtitlers attribute the frequent conversion of *gelijk* into *als* to spatial and temporal restrictions. About *efkes* and *seffens*, on the other hand, there has been a lot of discussion in the past. Initially, both lexemes were systematically converted into standard language, because the subtitlers supposed that *efkes* and *seffens* were dialect. At a certain moment, it turned out that Van Dale dictionary labels these words as 'colloquial' (BE, *spreektaal*) and, in accordance with their guidelines, CBD words should be reproduced in the subtitles. Both subtitlers, however, agree that *efkes* and *seffens* 'feel more dialectical compared to *kuisvrouw* and *ajuin*', which triggers them to use the BSD variant. Especially *efkes* has a strong dialectical connotation, because of the *-ke* diminutive it contains. Based on the interviews, it turns out that *efkes* as well as *seffens* are two delicate questions, about which the subtitlers did not yet reach an agreement. Another remarkable lexeme is the colloquial variant *schoon*. Most of the time, *schoon* is transferred to the subtitles, except for once:

9. Goh *wete*, ik vond *da* toen een heel **schoon** kind, maar als ik nu foto's zie, dan denk ik van: oh, zo **schoon** was *dieje* precies toch *nie*. (Echt niet ok!)

Ik vond dat toen een **mooi** kind, maar als ik nu foto's zie, denk ik: zo **schoon** was die toch niet.

At first, I thought it was a **pretty** baby, but when I see pictures now, I think: well, in fact he wasn't that **pretty**.

Subtitled 1 declares that *schoon* has a similar connotation as *efkes* and *seffens*: 'In the subtitlers' mind, this word sounds very Flemish, which results in a reflex to "clean it up" by translating it into standard language'. She admits that she would use *schoon* in a subtitle, but only because she knows that Van Dale dictionary labels it as CBD. Subtitled 2 adds to this point that consistency in subtitling is very important. If a subtitler opts for the colloquial variant the one time, he also has to use it the other, and vice versa. Usually, 'errors' similar to example 9 are eliminated during the final editing.

Contrary to their tolerant attitude toward the use of colloquial lexicon in television subtitling, both subtitlers declare that VRT is very strict toward morphosyntactic accuracy: 'In spoken language, a lot of errors are made against grammatical constructions and the guidelines prescribe that these "blunders" must be corrected'. Nevertheless, our data have shown that morphological colloquialisms (cf. Table 3) are occasionally reproduced in the subtitles, and syntactic colloquialisms (cf. Table 4) even quite often (more than 50% of the cases). The corpus examples below illustrate the occurrence of some of these morphological colloquial features in question.

10. Ja, *ma*, ik ben *ekik* ook *nie* verliefd op *u*, hè. (Thuis)
Ik ben ook niet verliefd op *u*.
I am not in love with *you* either.
11. Wij houden *u* gevangen. (Helden)
Wij houden *u* gevangen.
We keep *you* in prison.
12. Lieveke, ik begrijp *da ge da* event wilt organiseren. (Thuis)
Lieveke, ik begrijp dat je dat event wil organiseren.
Darling, I understand you want to organize this event.
13. Euh ja, dan hebben wij niks te doen hè **make**, of wel? (Tom & Harry)
Ja, euh... Dan hebben wij niks te doen, hè **make**? Of wel?
Well uhm, we have nothing planned then, have we **mother**?

The reproduction of the informal object *u* in the examples 10 and 11 has no specific function, according to the interviewed subtitlers. In theory,

the colloquial *u*-form is systematically replaced by the BSD *je/jou*-form. However, it occurs on occasion that subtitlers unconsciously copy the *u*-form to the subtitles. This can be attributed to the existence of a formal *u(w)* in the standard language, which is the polite form to address the second person singular (e.g. *Meneer, u hoeft zich geen zorgen te maken over uw bagage* ‘Sir, you don’t have to worry about your luggage’). As a consequence, subtitlers do often not realize that they are using the informal, colloquial variant. It is one of the most frequent mistakes and even during the final editing, editors easily miss it, especially when there is more than one error in the text. With regard to the *-ke* diminutive, subtitler 1 explains that in T888-subtitling this colloquial morpheme is exceptionally allowed when referring to nicknames or pet names. Especially in fiction programs, characters use these diminutives when speaking to their family or loved ones (e.g. *schatteke, lieveke, bolleke* ‘darling’, *moeke* ‘mommy’, *pake* ‘daddy’). In this context, the Flemish television soap *Thuis* took a pioneer role.⁴ Some characters in the soap are systematically addressed by a diminutive (e.g. *Simonneke* ‘Simonne’, *make* ‘mommy’, *Lowieke* ‘Lowie’) that is almost considered as their real name, because it is exclusively used to refer to one specific person. For this reason, VRT also started to use these nicknames in the subtitles of the soap. Through the years, the use of this *-ke* diminutive in nicknames and pet names also became common practice in the subtitles of other fiction programs.

The quantitative analyses in 4.1.1. did not only point out that morphological colloquialisms are occasionally reproduced in the subtitles, but also a couple of colloquial syntactic features are copied to the subtitles (cf. some corpus examples below).

14. Euh, **da moet** ik zelfs **nie** opzoeken. (Tom & Harry)
 Dat **moet** ik zelfs **niet** opzoeken.
 I **don’t have** to look it up.
15. En die hogere risicopremie, die zal **moeten betaald worden**. (De vrije markt)
 Die risicopremie zal **moeten betaald worden**.
 This insurance premium **must be paid**.

Both subtitlers 1 and 2 declare that the ‘errors’ in example 14 and 15 should have been corrected. VRT is very strict with regard to grammatical correctness, not only in spoken language use, but also in subtitling, and subtitlers are not free to choose whether they copy a colloquial construction or not. Nowadays, subtitlers regularly receive feedback concerning some prevalent

syntactic constructions that are often used erroneously. If the editors notice that a subtitler systematically makes the same ‘mistake’, an e-mail with feedback is sent to the entire team by the head of the department, who had been involved in the development of VRT’s subtitling guidelines. Subtitler 1 explains that *niet moeten* only recently came to the attention and the frequent use of this colloquial feature in our corpus proves that at least some subtitlers are still not aware that *niet moeten* is a colloquial construction. According to subtitler 1, *niet moeten* (‘not have to’) and the verbal end group *moeten betaald worden* (‘must be paid’) must have been reproduced by accident. The reproduction of *moeten betaald worden* in the TV program *De vrije markt* can be explained by the pressure of time. *De vrije markt* is always subtitled last-minute, because the subtitling department receives this program only a couple of hours before it is broadcast. As a result, there is no final editing, so ‘inconspicuous’ grammatical colloquialisms are not always removed from the subtitles because of this lack of time. On the other hand, subtitlers get plenty of time to prepare the subtitles of fiction series like *Tom & Harry*, so the reproduction of the colloquial feature *niet moeten* in this TV program cannot be attributed to lack of time. According to subtitler 1, few subtitlers initially knew that this construction is not correct. As no attention was given in the past to the use of *niet moeten*, it often appeared in the subtitles.

In sum, the interviews have given us a valuable insight into the linguistic choices that subtitlers make when subtitling a television program. First, it was repeatedly shown that the subtitlers attach great value to the normative advice of the Dutch language authorities on which the subtitling guidelines of the public broadcaster are based. For instance, both subtitlers mentioned that they regularly consult Van Dale dictionary during the subtitling process. Since VRT’s subtitling guidelines prescribe that colloquial lexicon should be reproduced in the subtitles, a subtitler will use a word that is labelled CBD, whereas dialect words will be translated into BSD. Furthermore, since VRT is very strict toward grammatical accuracy, *Taaladvies* is frequently consulted to ensure that all morphosyntactic colloquialisms are corrected and converted into BSD in the subtitles. Secondly, the interviews revealed that subtitlers assign different levels of colloquiality to the individual colloquialisms. These connotation differences are also implied in the terminology they use to refer to the colloquialisms. Although Van Dale dictionary labels the following words equally as CBD, the subtitlers declared that, for instance, *efkes*, *seffens*, *schoon*, and *ambras* ‘feel more dialectical’ compared to *ajuin* and *kuisvrouw*. They attribute this difference to the common use of *ajuin* and *kuisvrouw* in everyday spoken and written

contexts. As a result, these two lexemes are supposed to be generally known by the Flemish audience and because of this general intelligibility, they are also used in television subtitling. On the contrary, *efkes*, *seffens*, and *schoon* are perceived too dialectal, so subtitlers generally convert these lexemes into BSD, thus ignoring the labelling of Van Dale and the instructions of the subtitling style guide. In addition, the subtitlers admitted that some colloquial features are difficult to detect, as a result of their 'frequent occurrence in everyday language'. In other words, the informal object *u*, the colloquial position of the participle in the verbal end group group and the construction *niet moeten*, for example, are three non-salient colloquial features that are often unconsciously copied to the subtitles. Finally, the interviews confirmed that the program genre determines the use of CBD or BSD to a certain extent. Although the subtitle guidelines do not explicitly prescribe these genre differences, the subtitlers insinuated several times that certain colloquial lexemes (e.g. *ambras* and *schoon*) and morphological features (e.g. *-ke* diminutive) are only used in fiction series, and especially in the soap *Thuis*, to retain the 'couleur locale'. In informative programs such as *De Zevende Dag*, these colloquialisms must be translated into BSD.

5 Concluding remarks

In light of the increasing use of CBD in spoken public media, this paper aimed to investigate how subtitlers in Flanders deal with linguistic variation in the spoken source text on the Flemish public broadcaster VRT. For many years, the growing popularity of *tussentaal* has been a source of great concern, which repeatedly resulted in heated discussions among Dutch linguists and politicians. However, in the past two decades, this evaluative and rather negative focus on *tussentaal* has evolved into empirical research, in which the use of this colloquial variety is investigated in various situational contexts and in which the future position of *tussentaal* is speculated. These studies have demonstrated that CBD has recently entered the public institutions such as the public broadcast media and education. Given the increasing use of this colloquial variety, not only in informal spoken contexts, but also in the spoken language of public institutions, this study verified whether CBD also infiltrates into the written language (viz. subtitles) of public authorities. Comparing the original speech of twenty Flemish television programs, broadcast by VRT between 2014 and 2016, to the corresponding intralingual subtitles, we analyzed the linguistic choices of subtitlers in five program genres. More specifically,

it was investigated to what extent the subtitlers reproduce the lexical, morphological, and syntactic colloquialisms of the spoken source text in the subtitles, and whether the program genre influences these linguistic choices. In addition, we interviewed two subtitlers and the head of VRT's subtitling department in order to receive their comments on the results of our corpus study.

First, it was revealed that CBD lexemes are frequently reproduced in subtitles on Flemish television, whereas colloquial morphological items are generally converted into BSD. Syntactic colloquialisms take up a middle position. The frequent reproduction of CBD lexicon can be explained by VRT's tolerance towards the use of colloquial lexemes in the subtitles. By doing so, the public broadcaster wants to meet the audience's wishes to repeat the spoken source text as literally as possible in the subtitles. The general conversion of morphological colloquialisms is also in accordance with VRT's subtitling guidelines. However, our results have demonstrated that the informal object *u* and the *-ke* diminutive are not systematically converted into BSD. On the one hand, the subtitlers explain this by the specific function of the *-ke* diminutive to refer to pet- and nicknames in fiction series. On the other hand, the informal object *u* is often unconsciously reproduced by the subtitlers, because they are not always aware of this colloquial variant. For similar reasons, syntactic colloquialisms are frequently used in the subtitles. According to the subtitlers, the colloquial position of the participle in the verbal end group and the construction *niet moeten*, for example, are difficult to detect, because these CBD variants are very commonly used constructions in everyday language. Based on these results, it can be said that certain morphological and syntactic colloquialisms seem to be highly entrenched into the subtitlers' cognitive language system. In other words, the use of these CBD variants 'has become a highly automated routine' (Schmid, 2007, p. 118). Several studies have associated this cognitive entrenchment to word frequency (e.g. Bardovi-Harlig, 1987; Bybee, 2001; Schmid, 2010). According to Langacker (1987, p. 51), 'every use of a structure has a positive impact on its degree of entrenchment, whereas extended periods of disuse have a negative impact'. Simultaneously, aforementioned studies presume that there is also a directly proportional relation between the frequency of a variant and its salience. In our study, however, the subtitlers themselves admitted that it is difficult to recognize commonly used colloquial variants, which contradicts the assumption that deeply entrenched linguistic features are cognitively salient features that 'have a better chance of entering our focus of attention' (Schmid, 2007, p. 120).

Second, both the quantitative analyses and the subtitlers themselves indicated that the program genre influences the linguistic choices that are made in the subtitles. Whereas the Belgian-Dutch colloquialisms are generally translated into standard language in the subtitles of informative genres (documentaries) and programs with an educational role (children's television), the number of colloquial features increased significantly in the subtitles of humorous (comedy) and entertainment (fiction and light entertainment) programs. Since the use of colloquial language in these programs is a conscious strategy to create an authentic, amusing or comic atmosphere, it does not surprise that subtitlers also want to create this spontaneous effect in the subtitles, thereby reproducing the informal linguistic items of the original footage. Although the subtitles of documentaries also contain a high amount of CBD lexicon, the low number of attestations reduces the representativeness of this result.

In general, it can be assumed that nowadays CBD or *tussentaal* is no longer exclusively used in the spoken registers, since it also occurs in the written language of the policymakers. As a consequence, the question arises whether this will effect a new shift in the current standard language ideology, which holds the belief that BSD is the only appropriate variety to use in formal contexts and in written language. At the beginning of this paper, it was asked whether BSD is losing its position as the one and only 'best language' and whether the status of *tussentaal* is upgrading from a merely informal spoken variety to a generally used, both spoken and written, informal language variety. In other words, is the language situation in Flanders characterized by a process of respectively destandardization or demotization? First of all, the results of our study seem to reflect the current standard language ideology at least partially. Especially in more formal and more informative written contexts, BSD continues to be the outstanding variety to be used. Nevertheless, the unconscious reproduction of certain colloquial features suggests that CBD is strongly entrenched into the Dutch language system, which makes us believe that the current status of certain CBD items should be reconsidered and, if necessary, should be changed into BSD, as professional language users also regard them as such.

In our view, it seems that the 'pool' with standard language features is being complemented with more colloquial features. In addition, language authorities promote the functionality of colloquial lexicon in more informal written contexts, because the use of standard language could damage the authentic character of the program. BSD is no longer the most appropriate language variety and even seems to have its

shortcomings in these contexts. In other words, the role of CBD has become increasingly important in Flanders, and especially in national broadcasting contexts, it is gaining a more central place. In this respect, our results seem to support the demotization hypothesis. A process of destandardization seems implausible, since the condition of a ‘crumbling standard language ideology’ is not fulfilled (Ghyselen et al., 2016, p.84). On the contrary, BSD is still believed to be the outstanding written variety, also in television subtitling. Nevertheless, if we want to draw watertight conclusions about the dynamic processes that affect the position of Dutch in Flanders, in-depth diachronic research is needed to substantiate these assumptions.

Notes

1. The first author holds a PhD fellowship with the Flemish Research Foundation (FWO).
2. *Taaladvies* is currently regarded as the main source of normative guidelines for language users in the Dutch language area.
3. The names of the interviewees will not be mentioned in order to preserve their privacy
4. *Thuis* (‘Home’) plays out the daily life of several middle class families and is broadcast each day by VRT since 1955.

Bibliography

- Assis Rosa, Alexandra, ‘Features of oral and written communication in subtitling’. *Benjamins Translation Library*, 34 (2001), pp. 213-222.
- Auer, Peter, ‘Dialect vs. standard: a typology of scenarios in Europe’. In: Bernd Kortmann & Johan VanDer Auwera (eds.). *The Languages and Linguistics of Europe. A comprehensive guide*. Berlin: Mouton De Gruyter, 2011, pp. 485-500.
- Auer, Peter, ‘Europe’s sociolinguistic unity, or: A typology of European dialect/standard constellations’. In: Nicole Delbecq, Johan Van Der Auwera & Dirk Geeraerts (eds.). *Perspectives on Variation: Sociolinguistic, historical, comparative*. Berlin/New York: Mouton De Gruyter, 2005, pp. 7-42.
- Bardovi-Harlig, Kathleen, ‘Markedness and salience in second-language acquisition’. *Language Learning* 37 (1987), pp. 385-407.
- Beheydt, Ludo, ‘De toekomst van het Nederlands. Een nieuwe taalstrijd?’. In: Ludo Beheydt et al. (eds.). *Tussen taal en staat: vooruitkijken voor Vlaanderen*. Leuven: Davidsfonds, 1993, pp. 75-94.
- Brannan, Matthew J. & Oultram, Teresa, ‘Participant Observation’. In: Gillian Symon & Catherine Cassell (eds.). *Qualitative organizational research: Core methods and current challenges*. London: Sage, 2012, pp. 296-313.
- Bybee, Joan, *Phonology and language use*, vol. 94. Cambridge: University Press, 2003.
- Coupland, Nikolas & Kristiansen, Tore, ‘SLICE: Critical perspectives on language (de)standardisation’. In: Nikolas Coupland & Tore Kristiansen (eds.). *Standard languages and language standards in a changing Europe*. Oslo: Novus Press, 2011, pp. 11-35.

- De Caluwe, Johan, 'Conflicting language conceptions within the Dutch speaking part of Belgium'. In: Rudolf Muhr (ed.). *Standardvariationen und Sprachideologien in verschiedenen Sprachkulturen der Welt/Standard Variations and Language Ideologies in Different Language Cultures around the World*. Peter Lang, 2003, pp. 53-58.
- De Caluwe, Johan, 'Tussentaal als natuurlijke omgangstaal in Vlaanderen'. *Studia Germanica Gandensia*, 2006, pp. 19-34.
- De Caluwe, Johan, 'Tussentaal wordt omgangstaal in Vlaanderen'. *Nederlandse taalkunde*, 14 (2009), pp. 8-25.
- De Decker, Benny, *De chattaal van Vlaamse tieners. Een taalgeografische analyse van Vlaamse (sub)standaardiseringsprocessen tegen de achtergrond van de internationale chatcultuur*. Antwerpen: University of Antwerp, 2014.
- De Ridder, Frederik, *Vlaamse tussentaal op televisie: een analyse van het fictieve reportageprogramma In de gloria*. Ghent: Ghent University, 2007.
- De Ridder, Reglindis, *Arme rijke taal! Audiovisual translation and minority language planning in pluricentric language areas: a case study of Flemish public service broadcasting*. Dublin: Dublin City University, 2015.
- De Schutter, Georges, 'Talen, taalgemeenschappen en taalnormen in Vlaams-België'. *Verslagen en Mededelingen van de Koninklijke Academie voor Nederlandse Taal- en Letterkunde 108* (1998), pp. 227-251.
- Delaere, Isabelle & De Sutter, Gert, 'Applying a multidimensional, register-sensitive approach to visualize normalization in translated and non-translated Dutch'. *Belgian Journal of Linguistics*, 27 (2013), pp. 43-60.
- Delaere, Isabelle, De Sutter, Gert & Plevoets, Koen, 'Is translated language more standardized than non-translated language?: Using profile-based correspondence analysis for measuring linguistic distances between language varieties'. *Target*, 24 (2012), pp. 203-224.
- Delarue, Steven, 'Contrasterende (standaard) taalideologieën bij Vlaamse leerkrachten: een Gentse casestudy'. *Handelingen: Koninklijke Zuid-Nederlandse Maatschappij voor Taal- en Letterkunde en Geschiedenis*, 67 (2014), pp. 219-248.
- Delarue, Steven, *Bridging the policy-practice gap: how Flemish teachers' standard language perceptions navigate between monovarietal policy and multivarietal practice*. Ghent: Ghent University, 2016.
- Díaz-Cintas, Jorge & Remael, Aline, *Audiovisual Translation, Subtitling*. London/New York: Routledge, 2014.
- Díaz-Cintas, Jorge, 'Subtitling'. In: Yves Gambier & Luc van Doorslaer (eds.) *Handbook of Translation Studies*. Amsterdam/Philadelphia: John Benjamins, 2010, pp. 344-349.
- Duranti, Alessandro, *Linguistic Anthropology*. Cambridge: University Press, 1997.
- Ericsson, K. Anders & Simon, Herbert A., *Protocol analysis (revised edition). Overview of Methodology of Protocol Analysis*. Cambridge: MIT Press, 1993.
- Ferguson, Charles A., 'Diglossia'. *Word* 15 (1959), pp. 325-340.
- Fishman, Joshua A., 'Bilingualism With and Without Diglossia; Diglossia With and Without Bilingualism'. *Journal of Social Issues* 23 (1967), pp. 29-38.
- Geeraerts, Dirk & Speelman, Dirk, 'Destandardization and demotization from a lectometrical perspective'. *Taal & Tongval Colloquium*, 2014.
- Geeraerts, Dirk, 'Hoe gans het volk is de taal?' *Over Taal* 38 (1999), pp. 30-34.
- Ghyselen, Anne-Sophie, 'Ne zelfzekere leraar of gewoon nen enthousiaste mens? Een matched-guise onderzoek naar de attitude tegenover tussentaal bij West-Vlamingen'. *Taal en Tongval* 61 (2009), pp. 83-113.
- Ghyselen, Anne-Sophie, Delarue, Steven & Lybaert, Chloé, 'Studying standard language dynamics in Europe'. *Taal en Tongval*, 68 (2016), pp. 75-91.

- Ghyselen, Anne-Sophie, *Verticale structuur en dynamiek van het gesproken Nederlands in Vlaanderen: een empirische studie in Ieper, Gent en Antwerpen*. Ghent: Ghent University, 2016.
- Goossens, Jan, 'De toekomst van het Nederlands in Vlaanderen'. *Ons Erfdeel* 43 (2000), pp. 2-13.
- Grondelaers, Stefan & Van Hout, Roeland, 'The standard language situation in the Low Countries: Top-down and bottom-up variations on a diaglossic theme'. *Journal of Germanic Linguistics*, 23 (2011), pp. 199-243.
- Grondelaers, Stefan, Van Hout, Roeland & Speelman, Dirk, 'A perceptual typology of standard language situations in the Low Countries'. *Standard languages and language standards in a changing Europe*, 2011, pp. 199-222.
- Hamaida, Lena, 'Subtitling slang and dialect'. *Proceedings of the MuTra 2007-LSP Translation Scenarios Conference*, 2007, pp. 1-11.
- Hendrickx, Ruud, 'Laat ons ne keer te goei naar onszelf luisteren'. *VRT. Taalsignaal voor 6 miljoen Vlamingen*. <<http://www.vrt.be/taal/laat-ons-ne-keer-te-goei-naar-onszelf-luisteren>>, 2001.
- Hendrickx, Ruud, 'Nederlands in het Nederlands ondertitelen'. *VRT-Taalnet*. <http://taal.vrt.be/nederlands-in-het-nederlands-ondertitelen>, 2011.
- Hendrickx, Ruud, 'Wat zegt ie?'. *VRT-Taalnet*. <<https://taal.vrt.be/wat-zegt-ie?print=1>>, 2003.
- Hendrickx, Ruud, *Taalcharter 2012*. <<https://taal.vrt.be/taalcharter-2012>>, 2012.
- Hendrickx, Ruud, *Taalcharter*. <<http://www.vrt.be/taal/taalcharter>>, 1998.
- Impe, Lenn & Speelman, Dirk, 'Vlamingen en hun (tussen)taal. Een attitudeel mixed-guise onderzoek'. *Handelingen van de Koninklijke Zuid-Nederlandse Maatschappij voor Taal- en Letterkunde en Geschiedenis* 61 (2007), 109-128.
- Jacobs, Geert, Van Hout, Tom & Van Praet, Ellen (eds.). 'Discursive Perspectives on News Production'. *Journal of Pragmatics* (2011). Amsterdam: Elsevier.
- Lefevere, Elien, *Tussentaal in de Vlaamse media. Een onderzoek naar het taalgebruik van televisiepresentatoren*. Ghent: Ghent University, 2011.
- Lopes Cavalheiro, Lili, 'Linguistic Variation in Subtitling for Audiovisual Media in Portugal'. *Linguistica Antverpiensia*, 7 (2008), pp. 17-28.
- Lybaert, Chloé, 'De ideologische positie van Algemeen Nederlands in Vlaanderen'. *Taal en Tongval*, 67 (2015), pp. 97-134.
- Lybaert, Chloé, 'Perceptie van tussentaal in het gesproken Nederlands in Vlaanderen'. *Nederlandse Taalkunde*, 19 (2014b), pp. 185-220.
- Lybaert, Chloé, *Het gesproken Nederlands in Vlaanderen: percepties en attitudes van een spraakmakende generatie*. Ghent: Ghent University, 2014a.
- Mattheier, Klaus J., 'Über destandardiseering, Umstandardisierung and Standardisierung in modernen Europäischen Standardsprachen'. In Klaus J. Mattheier & Edgar Radtke (eds.): *Standardisierung und Destandardisierung Europäischer Nationalsprachen*, 1-9. Frankfurt: Peter Lang Verlag, 1997.
- McIlvenny, Paul, Mettovaara, Sari & Tapio, Ritva, 'I really wanna make you laugh: stand-up comedy and audience response'. *Folia, Fennistica and Linguistica: Proceedings of the Annual Finnish Linguistic Symposium May*, 1992, pp. 225-247.
- Myers, Fred R., *Pintupi Country Pintupi Self: Sentiment, Place, and Politics among Western Desert Aborigines*. Washington: Smithsonian Institution Press, 1986.
- Nikken, P. & Friebe, C., *Tussen Kijkcijfers en Boodschap. Omroepmedewerkers over Jeugdprogrammering*. Leiden: Rijksuniversiteit Leiden, 1990.
- Plevoets, Koen, 'De status van de Vlaamse tussentaal: een analyse van enkele socio-economische determinanten'. *Tijdschrift voor Nederlandse Taal- en Letterkunde*, 129 (2013), pp. 191-233.
- Plevoets, Koen, *Tussen spreek-en standaardtaal: een corpusgebaseerd onderzoek naar de situationele, regionale en sociale verspreiding van enkele morfo-syntactische verschijnselen uit het gesproken Belgisch-Nederlands*. Leuven: Katholieke Universiteit Leuven, 2008.

- Prieels, Lynn & De Sutter, Gert, 'In-between language policy and language reality: a corpus-based multivariate study of the interlingual and intralingual subtitling practice in Flanders'. *Perspectives: Studies in Translatology*, 2017.
- Prieels, Lynn, 'Tussentaal in tv-programma's'. In: Johan De Caluwe, Steven Delarue, Anne-Sophie Ghyselen, & Chloé Lybaert (eds.), *Tussentaal: over de talige ruimte tussen dialect en standaardtaal in Vlaanderen*. Gent: Academia Press, 2013, pp. 35-49.
- Prieels, Lynn, Delaere, Isabelle, Plevoets, Koen & De Sutter, Gert, 'A corpus-based multivariate analysis of linguistic norm-adherence in audiovisual and written translation'. *Across Languages and Cultures*, 16 (2015), pp. 209-231.
- Ramos Pinto, Sara, 'How important is the way you say it?: A discussion on the translation of linguistic varieties'. *Target* 21 (2009), pp. 289-307.
- Remael, Aline, 'Mainstream Narrative Film Dialogue and Subtitling'. *The Translator*, 9 (2003), pp. 225-247.
- Remael, Aline, De Houwer, Annick & Vandekerckhove, Reinhild, 'Intralingual open subtitling in Flanders: Audiovisual translation, linguistic variation and audience needs'. *The Journal of Specialised Translation*, 10 (2008), pp. 76-105.
- Rutter, Jason, *Stand-up as interaction: Performance and audience in comedy venues*. Salford: University of Salford, 1997.
- Rys, Kathy & Taeldeman, Johan, 'Fonologische ingrediënten van Vlaamse tussentaal'. In: Dominiek Sandra, Rita Rymenans, Pol Cuvelier & Peter Van Petegem (eds.), *Tussen taal, spelling en onderwijs. Essays bij het emeritaat van Frans Daems*. Ghent: Academia Press, 2007, pp. 1-8.
- Schmid, Hans-Jörg, 'Does frequency in text instantiate entrenchment in the cognitive system'. *Quantitative methods in cognitive semantics: Corpus-driven approaches*, 2010, pp. 101-133.
- Schmid, Hans-Jörg, 'Entrenchment, salience, and basic levels'. In: Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford handbook of cognitive linguistics*. Oxford University Press, 2007.
- Selinker, Larry, 'Interlanguage'. *IRAL-International Review of Applied Linguistics in Language Teaching*, 10 (1972), pp. 209-232.
- Sleurs, Kim & Jacobs, Geert, 'Beyond preformulation: an ethnographic perspective on press releases'. *Journal of Pragmatics*, 37 (2005), pp. 1251-1273.
- Speelman, Dirk, Grondelaers, Stefan & Geeraerts, Dirk, 'Profile-based linguistic uniformity as a generic method for comparing language varieties'. *Computers and the Humanities*, 37 (2003), pp. 317-337.
- Taeldeman, Johan, 'Welk Nederlands voor Vlamingen?' *Nederlands van Nu* 40 (1992), pp. 33-51.
- Van Bree, Cor, 'Taalbewustzijn, taalverandering en regionale variatie. Een benadering'. *Taal en Tongval* 52 (2000): pp. 22-46.
- van Dale, Johan Hendrik, den Boon, Ton, Hendrickx, Ruud & Van der Sijs, Noline, *Van Dale: Groot woordenboek van de Nederlandse taal: 3DI*, 15^e editie. Utrecht: Van Dale Uitgevers, 2015.
- Van de Velde, Hans, 'Verkavelingsvlaams. Wat is mij dat nu?'. In Roeland Van Hout & Joep Kruijssen (eds.), *Taalvariaties. Toonsettingen en modulaties op een thema*. Dordrecht: Foris Publications, 1996, pp. 261-271.
- Van Gijssel, Sofie, Speelman, Dirk & Geeraerts, Dirk, 'Style shifting in commercials'. *Journal of pragmatics*, 40 (2008), pp. 205-226.
- Van Hoof, Sarah, 'Feiten en fictie-Taalvariatie in Vlaamse televisiereeksen vroeger en nu'. *Nederlandse Taalkunde*, 18 (2013a), pp. 35-64.
- Van Hoof, Sarah, *Feiten en fictie: een sociolinguïstische analyse van het taalgebruik in fictieseries op de Vlaamse openbare omroep (1977-2012)*. Antwerpen: Universiteit Antwerpen, 2013b.
- Van Istendael, Geert, *Het Belgisch labyrint: de schoonheid der wanstaltigheid*. Amsterdam: Arbeiderspers, 1989.

- Vancompernelle, Hélène, *Normgevoeligheid: attitude van Vlaamse jongeren ten aanzien van het Standaardnederlands, de tussentaal en het dialect*, vol. 5. Ghent: Academia Press, 2012.
- Vandekerckhove, Reinhild, 'Tussentaal as a source of change from below in Belgian Dutch. A case study of standardization processes in the chat language of Flemish teenagers'. *Germanic language histories "from below" (1700-2000)*. Berlin: De Gruyter, 2007, pp. 189-203.
- Vandekerckhove, Reinhild, De Houwer, Annick & Remael, Aline, 'Intralinguale ondertiteling op de Vlaamse televisie: een spiegel voor de taalverhoudingen in Vlaanderen'. *Tussen Taal, Spelling en Onderwijs: Essays bij het Emeritaat van Frans Daems*, 2007, pp. 71-83.
- Vandekerckhove, Reinhild, De Houwer, Annick, Remael, Aline & Van der Niepen, Isabelle, 'Intralinguale ondertiteling van Nederlandstalige televisieprogramma's in Vlaanderen: Linguïstische en extra-linguïstische determinanten'. In *Artikelen van de vijfde sociolinguïstische conferentie*, 2006, pp. 503-513.
- VRT Vertalingen ondertiteling, *Stijlboek 888: Teletekstondertiteling* [internal document] Brussel: VRT-teletekstredactie, 2009.
- Wayne Langacker, Ronald, *Foundations of Cognitive Grammar. Theoretical prerequisites* (Vol. 1). Stanford: Stanford University Press, 1987.
- Zenner, Eline, Geeraerts, Dirk & Speelman, Dirk, 'Expeditie Tussentaal-Leef tijd, identiteit en context in "Expeditie Robinson"'. *Nederlandse taalkunde*, 14 (2009), pp. 26-44.

About the authors

Lynn Prieels, Ghent University, Department of Translation, Interpreting and Communication, Groot-Brittanniëlaan 45, 9000 Ghent, Belgium.
E-mail: Lynn.Prieels@UGent.be

Gert De Sutter, Ghent University, Department of Translation, Interpreting and Communication, Groot-Brittanniëlaan 45, 9000 Ghent, Belgium.
E-mail: Gert.DeSutter@UGent.be

Appendices

Appendix 1 Corpus examples of each of the linguistic features

A. Lexical features

- Feature 1: *accident*
traffic accident
Allee, ik heb precies een accident gehad.
'Gee, it is like I had an accident.'
- Feature 2: *ajuin*
- Translation: *onion*

- Corpus example: *Onze papa deed dat ook met ajuin.*
'Our dad did that with onion too.'
- Feature 3: *ambras*
Translation: *quarrel*
- Corpus example: *Hoe? Is ze ambras komen maken?*
'What? Did she come here to pick a quarrel?'
- Feature 4: *appelsien*
Translation: *orange*
- Corpus example: *Pelé à vif, dat is een beetje zoals een appelsien.*
'Pelé a vif, it is a bit like an orange.'
- Feature 5: *boeleke*
Translation: *pet name for a baby*
- Corpus example: *Waar is dat klein boeleke hier?*
'Where is the little baby?'
- Feature 6: *afbollen*
Translation: *get out*
- Corpus example: *Bol het af, jong.*
'Get out, you.'
- Feature 7: *bról*
Translation: *trash*
- Corpus example: *600 euro aan bról die we niet nodig hebben.*
'600 euros of trash that we don't need.'
- Feature 8: *buizen*
Translation: *to flunk*
- Corpus example: *Maar je bent gebuisd voor lo, Kleine.*
'But you flunked PE, little one.'
- Feature 9: *camionette*
Translation: *delivery van*
- Corpus example: *Er was eens iets met de motor van onze pa zijn camionette.*
'One day, something went wrong with our father's delivery van.'
- Feature 10: *chance*
Translation: *luck*
- Corpus example: *Chance dat dat niet gevallen is.*
'Lucky it didn't fall!'
- Feature 11: *chapelure*
Translation: *breadcrumbs*
- Corpus example: *Daar gaan we een eitje onder pletten, chapelure, wat bijkruiden.*
'Then we add a crushed egg, some breadcrumbs, some spices.'
- Feature 12: *chichi madam*
Translation: *chichi lady (negative connotation)*
- Corpus example: *Maar voor die chichi madam wil ik niet meer werken.* 'But I don't want to work anymore for that chichi lady.'
- Feature 13: *dagdagelijks*
Translation: *daily*

- Corpus example: ..die wij op school vaak op dagdagelijkse basis meemaken.
'... which happens daily at school.'
- Feature 14: *efkes*
Translation: *just (temporal)*
Corpus example: *Heel efkes....*
'Just a minute...'
- Feature 15: *flik*
Translation: *cop*
Corpus example: *Daar, de flikken.*
'There are the cops.'
- Feature 16: *fretten*
Translation: *to scoff (food)*
Corpus example: *Ik fret chips.*
'I scoff chips.'
- Feature 17: *frigo*
Translation: *fridge*
Corpus example: *Het kan gaan over het krediet van je wagen, je frigo of je huis.*
'It can be about the credit of your car, your fridge or your home.'
- Feature 18: *in het hol van Pluto*
Translation: *at the back of beyond*
Corpus example: *Ik ga niet afspreken in het hol van Pluto.*
'I am not going to meet at the back of beyond.'
- Feature 19: *gelijk*
Translation: *like (comparison)*
Corpus example: *Gelijk Paulien en Ruben.*
'Like Paulien and Ruben.'
- Feature 20: *kostelijke affaire*
Translation: *expensive deal*
Corpus example: *Goh, die parking hier, zeg. Kostelijke affaire, hoor.*
'Phew, that car park is an expensive deal, isn't it.'
- Feature 21: *kozijn*
Translation: *cousin*
Corpus example: *Onze pa laat zijn kozijn daarnaar kijken.*
'Our dad will show it to his cousin.'
- Feature 22: *kuisen*
Translation: *to clean*
Corpus example: *Om het huis te kuisen en te koken.*
'To clean the house and to cook.'
- Feature 23: *kuisvrouw*
Translation: *cleaning lady*
Corpus example: *Een kuisvrouw kost geld, alles kost geld tegenwoordig.*
'A cleaning lady costs money, everything costs money nowadays.'
- Feature 24: *madam*
Translation: *madam*

Corpus example:	<i>Maar madam, jij moet je toch niet excuseren.</i> 'But madam, you don't have to apologize.'
Feature 25:	<i>nonkel</i>
Translation:	<i>uncle</i>
Corpus example:	<i>Dat is het nummer van mijn nonkel.</i> 'That is the number of my uncle.'
Feature 26:	<i>omwille van</i>
Translation:	<i>because of</i>
Corpus example:	<i>...niet zozeer omwille van een gebrek aan startkapitaal.</i> '...not so much because of a lack of starting capital.'
Feature 27:	<i>patat</i>
Translation:	<i>patato</i>
Corpus example:	<i>Een patat zonder zout. . .</i> 'A patato without salt...'
Feature 28:	<i>plezant</i>
Translation:	<i>cheerful</i>
Corpus example:	<i>Wobbe, pépé, ik vond het superplezant.</i> 'Wobbe, gramps, I had so much fun.'
Feature 29:	<i>saucisse</i>
Translation:	<i>sausage</i>
Corpus example:	<i>En dat is met stukjes saucisse erin, ik denk Boulogne.</i> 'There are some pieces of sausage in it, I think it is Boulogne.'
Feature 30:	<i>schoon</i>
Translation:	<i>good-looking</i>
Corpus example:	<i>Zo schoon was die toch niet.</i> 'He was not that good-looking.'
Feature 31:	<i>seffens</i>
Translation:	<i>later</i>
Corpus example:	<i>En dan kunnen we seffens, als ze in de oven steken.</i> 'And later, when they are in the oven, we can'
Feature 32:	<i>sjotten</i>
Translation:	<i>play soccer</i>
Corpus example:	<i>Mijn vlees is aan het rusten. Sjotten?</i> 'The meat is resting. Let's play soccer'
Feature 33:	<i>smossen</i>
Translation:	<i>to make a mess of</i>
Corpus example:	<i>Ben je aan het smossen?</i> 'Are you making a mess of it?'
Feature 34:	<i>stoefen</i>
Translation:	<i>to brag</i>
Corpus example:	<i>Ik ben dat gestoef van Toon beu.</i> 'I am tired of Toon's bragging.'
Feature 35:	<i>vijzen</i>
Translation:	<i>to screw</i>

- Corpus example: *Kom, vijs die deur aan de kast en hou je bakkes.*
'Come on, screw this door onto the closet and shut up.'
- Feature 36: *weeral*
Translation: *again*
Corpus example: *Wat? Weeral?*
'What? Again?'
- Feature 37: *eens*
Translation: *as soon as*
Corpus example: *Eens die basisbehoeften vervuld zijn..*
'As soon as these basic needs are fulfilled..'
- Feature 38: *zot*
Translation: *crazy*
Corpus example: *Zot, ik ben zes keer gebuisd hè.*
'Are you crazy? I flunked six times.'
- Feature 39: *zever*
Translation: *twaddle*
Corpus example: *Och, negen van de tien is het zever wat de mensen zeggen.*
'Oh, nine out of ten people are twaddling.'
- Feature 40: *zwanzen*
Translation: *to joke*
Corpus example: *Ik zwans maar, hè.*
'Oh, I am joking.'

B. Morphological features

- Feature 1: *adjectief (+e)*
Translation: *colloquial flexion of the adjective*
Corpus example: *Groot-Brittannië heeft het voordeel van een lagere pond als export..*
'Great Britain has the advantage of a lower pound as export..'
- Feature 2: *bezittelijk voornaamwoord (+e)*
Translation: *colloquial flexion of the possessive pronoun*
Corpus example: *Onze jubilee, dat wordt een ramp, hè.*
'Our anniversary is going to be a disaster, isn't it?'
- Feature 3: *diminutief -ke*
Translation: *diminutive*
Corpus example: *Pake, voorzichtig.*
'Daddy, be careful.'
- Feature 4: *ikke*
Translation: *flexion of the personal pronoun I*
Corpus example: *Ikke Samson.*
'I [want] Samson'
- Feature 5: *object u*
Translation: *object you*
Corpus example: *Ik ben ook niet verliefd op u.*
'I am not in love with you either.'

C. Syntactic features

Feature 1:	comparatief + <i>dan</i> + object
Translation:	comparative
Corpus example:	<i>Ik heb meer dan jou.</i> 'I have more than you.
Feature 2:	<i>durven</i> + inf
Translation:	<i>to dare</i> + infinitive
Corpus example:	<i>We durven uw event niet organiseren, we zijn nog niet klaar.</i> 'We don't dare to organize your event, we are not ready yet.'
Feature 3:	<i>zijn ontslag geven</i>
Translation:	<i>to resign</i>
Corpus example:	<i>Ja, daarom heb ik mijn ontslag gegeven.</i> 'Yes, that is why I have resigned.'
Feature 4:	<i>niet moeten</i>
Translation:	<i>not have to</i>
Corpus example:	<i>Je moet niet bang zijn.</i> 'You don't have to be afraid.'
Feature 5:	vz + vz
Translation:	preposition + preposition
Corpus example:	<i>..en probeert zich niet naar voor te schuiven als het machtsblok.</i> '..and doesn't try to move forwards as power block.'
Feature 6:	<i>zijn</i>
Translation:	possessive <i>zijn</i>
Corpus example:	<i>Nathalie, mag je iemand anders zijn kind straffen?</i> 'Nathalie, is one allowed to punish someone else's child?'
Feature 7:	<i>onze/ons</i> + eigennaam/soortnaam
Translation:	<i>our</i> + proper name/generic name
Corpus example:	<i>Ja, ik ben hier met mijn man en onze Harry en Babette.</i> 'Yes, I am here with my husband and Harry and Babette.'
Feature 8:	<i>de</i> + eigennaam
Translation:	<i>the</i> + proper name
Corpus example:	<i>Dit is de Frans.</i> 'This is Frans.'
Feature 9:	<i>zet je erbij</i>
Translation:	<i>have a seat</i>
Corpus example:	<i>Allee H�el�ene, zet je erbij.</i> 'Come on, H�el�ene, have a seat.'
Feature 10:	aux + part + inf
Translation:	position of the participle in the verbal end group
Corpus example:	<i>Die risicopremie zal moeten betaald worden.</i> 'This insurance premium must be paid.'

D. Additional features

Feature 1:	<i>verbuiging lidwoord: den, ne(n)</i>
Translation:	flexion of the article
Corpus example:	<i>Hij zal ne fantastische papa zijn.</i> 'He will be a great dad.'
Feature 2:	<i>verbuiging aanwijzend vnv: diene(n), dieje(n), dezen</i>
Translation:	flexion demonstrative pronoun
Corpus example:	<i>Fantastisch nummer trouwens, diene Hold Back The River.</i> 'Great song, that Hold Back The River.'
Feature 3:	<i>persoonlijk voornaamwoord ge/gij</i>
Translation:	personal pronoun <i>you</i>
Corpus example:	<i>Ik weet nie of ge mij kunt horen, oma.</i> 'I don't know whether you can hear me, grandma.'
Feature 4:	<i>1ste persoon enkelvoud + n</i>
Translation:	colloquial conjugation 1st person singular
Corpus example:	<i>Ik zen gelukkig, hè.</i> 'Well, I am happy.'
Feature 5:	<i>2e persoon enkelvoud</i>
Translation:	colloquial conjugation 2nd person singular
Corpus example:	<i>Gij zijt 't probleem.</i> 'You are the problem.'
Feature 6:	<i>van/voor + beknopte bijzin</i>
Translation:	reduced clause
Corpus example:	<i>Deze sla is te goed voor te versnijden.</i> 'This lettuce is too good for cutting.'
Feature 7:	<i>dubbele negatie</i>
Translation:	double negative
Corpus example:	<i>Da had ik nooit nie gedacht da ik da ging winnen.</i> 'I never thought I would win this game.'
Feature 8:	<i>redundant dat</i>
Translation:	redundant <i>that</i>
Corpus example:	<i>Wa da gij voor de kinderen doet, dat is onbetaalbaar.</i> 'The things you do for the children cannot be repaid.'
Feature 9:	<i>subjectsreduPLICATIE</i>
Translation:	reduplication of the subject
Corpus example:	<i>Ja, ge moogt gij de creativiteit van de jeugd ook nie aan banden leggen, hè.</i> 'You must not curb the creativity of youth.'

Appendix 2 The total number of BSD and CBD attestations per dataset

Dataset	Label	Program genre									
		Light entertainment		Fiction		Comedy		Documentaries		Childrens' television	
			%		%		%		%		%
Lexical features	CBD	24	27.0%	25	25.8%	34	40.5%	4	4.5%	2	2.2%
	BSD	17	33.3%	9	17.6%	16	31.4%	1	2.0%	8	15.7%
Syntactic features	CBD	4	10.7%	13	39.3%	3	25.0%	4	7.1%	2	17.9%
	BSD	2	12.5%	10	45.8%	1	4.2%	6	4.2%	7	33.3%
Morphological features	CBD	2	7.1%	9	39.3%	10	35.7%	2	10.7%	2	7.1%
	BSD	71	21.5%	124	37.6%	90	27.3%	12	3.6%	33	10.0%

Appendix 3 Diagram with the normalized frequencies of the CBD and BSD variants in each feature set

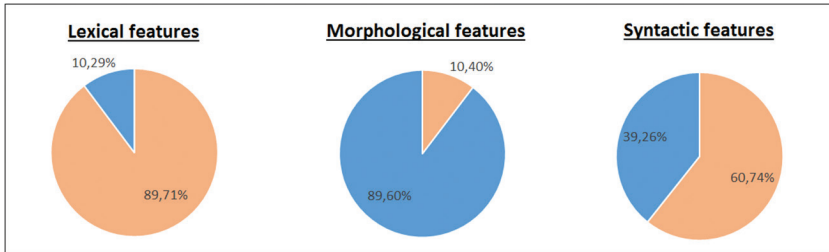


Figure 2 Normalized frequencies of the CBD (light grey) and BSD (dark grey) variants in each feature set