Nederlandse vertaling:
Balanceren tussen taalbeleid en taalrealiteit. Een corpusgebaseerde, multivariate study naar linguïstisch normconformisme in Belgisch-Nederlandse ondertiteling

Balancing between language policy and language reality

A corpus-based multivariate study on linguistic norm adherence in Belgian-Dutch subtitling

Lynn Prieels

Dankwoord

Walk on, walk on
With hope in your heart
And you'll never walk alone
(Gerry & The Pacemakers, 1963)

Het is moeilijk om te beschrijven welke gevoelens er door mij heen gaan wanneer ik voor de zoveelste maal door dit proefschrift blader. Opluchting, trots en voldoening zijn er maar enkele die spontaan bij me opkomen. Evenals een gevoel van verwondering dat vier jaren van hard labeur samengebundeld kunnen zitten in dit ene boek. Terwijl ik vol trots dit proefschrift in de handen houd, realiseer ik mij dan ook dat enkele steunende collega's, naaste familieleden en lieve vrienden een laatste woordje van dank verdienen.

In de eerste plaats wil ik mijn grootste dank betuigen aan mijn promotor, Gert De Sutter. Bedankt, Gert, om mij de afgelopen jaren te begeleiden in mijn onderzoeksproces. Van bij mijn projectaanvraag tot aan de laatste letter die ik hier heb neergeschreven ben jij mij blijven motiveren en gaf je mij dat bemoedigende schouderklopje als ik dat nodig had. Jouw inzichtrijke tips en constructieve feedback hebben ervoor gezorgd dat ik vandaag dit mooie proefschrift kan neerleggen.

Sarah, jouw sociolinguïstische expertise heeft een grote meerwaarde betekend voor mijn onderzoek. Als co-promotor stond jij voor mij klaar met verrijkende kennis en handige tips. Ook het gevoel van jouw aanwezigheid en bereidwilligheid om mij te helpen als dat nodig was, heeft mij tot aan het einde van dit project veel steun gegeven.

Bedankt, Koen Plevoets, Aline Remael, Reinhild Vandekerckhove en Anna Vermeulen, om deel geweest te zijn van mijn Doctoraatsbegeleidingscommissie. Jullie advies, suggesties en ook de tijd die jullie hebben geïnvesteerd in het begeleiden van mijn project, heb ik

Maar ook buiten de afdeling Nederlands hebben enkele fantastische collega's van de vakgroep VTC een inspirerende en boeiende werkplek gemaakt.

Céline, wij zaten slechts een verdieping van elkaar verwijderd – jij beneden bij Engels en ik boven bij Nederlands – en de weg naar een *goe* glas cava bracht ons dan ook maar al te vaak samen. Wat begon als goed studiegezelschap aan de Blandijn is uitgegroeid tot een mooie vriendschap die ik vandaag enorm koester. Ik hoop dan ook oprecht dat we, samen met onze wederhelften, nog veel gastronomische ervaringen zullen mogen delen, en vooral dat ik jou zal kunnen blijven bereiken op het nummer *wine one one*.

Bart en Cédric, bedankt om mij in te wijden in de wondere wereld van academische conferenties. Mijn eerste congres in Málaga was er één om nooit meer te vergeten. Katrien, Ann en Chantal, Ann en Sofie, jullie administratieve deskundigheid hebben ervoor gezorgd dat ik steeds de weg vond door het papierwerk en de strikte regels van het FWO. Daar ben ik jullie ontzettend dankbaar voor. Peter en Sam, talloze onderzoekspresentaties zouden in het honderd zijn gelopen als jullie er niet waren geweest met jullie kostbare IT-kennis. Verder wil ik ook nog enkele collega's bedanken die zorgden voor een gezellige sfeer tijdens de lunchpauzes. Bedankt, Bart en Bart, Bernard, Cynthia, Griet en Griet, Hinde, Linda, Nadine, Orphée, Sabine, Sofie, Veronique voor de luchtige babbels over uiteenlopende thema's. Van de liefde voor Hellofresh maaltijden tot het opduiken van enkele bizarre allergieën, het passeerde allemaal wel eens de docentenruimte.

Ook mijn nieuwe collega's aan de Universiteit van Bergen verdienen een woordje van dank. Zij hebben mij immers de ruimte gegeven om de laatste hand aan dit proefschrift te kunnen leggen. Carola, Gudrun, Elodie, Hanne, Inge, Rita, bedankt voor de morele steun en jullie begrip in een van de meest chaotische periodes uit mijn leven!

Wat ik zeker niet zal vergeten, is dat ik ook buiten de universiteit een leger aan vriendinnen heb die steeds voor mij klaarstaan en die me de afgelopen jaren door dik en dun hebben gesteund. Dankjewel, lieve Kellybellie, om die speciale vriendin in mijn leven te zijn. Jij gaf mij telkens weer het gevoel dat ik *ferm goe* bezig was en dat ik trots mag zijn op wat ik daarmee heb bereikt. Mijn girlies, onze talrijke dates zorgden meermaals voor het nodige verzetje tijdens stresserende en emotionele momenten. Het voelt enorm goed om deel te mogen uitmaken van zo'n hechte vriendengroep. Lieve Blandina's, als een bende gelijkgestemde taalkundigen hebben we niet alleen samen onze opleiding aan de Blandijn overleefd, maar onze vriendschap is ook in de jaren ná onze Blandijnperiode blijven standhouden. Bedankt voor de gezellige babbels, heerlijke brunches en etentjes, zotte feestjes en vooral om er voor mij te zijn. Laura, Jolien en Evelien, ook met jullie enthousiasme, steunende woorden en gekke persoonlijkheden hebben jullie de stress meermaals kunnen relativeren.

Een speciale vermelding gaat uiteraard naar mijn familie. In de eerste plaats ben ik mijn ouders enorm dankbaar voor de kansen die ik van kinds af van hen heb gekregen. Jullie onvoorwaardelijke steun heeft mij gestimuleerd om mijn dromen na te jagen en nooit op te geven. Een gemeende dankjewel gaat ook naar mijn schoonouders, (schoon)zus en schoonbroers omdat jullie op de juiste momenten polsten naar hoe het met mij (en 'het doctoraat') ging, ook al is het misschien vandaag nog steeds een groot mysterie voor jullie waar ik nu de afgelopen vier jaar precies mee bezig ben geweest. Een extra bedankje gaat naar mijn schoonbroer Glenn, die met zijn grafisch talent een grote hulp is geweest bij het ontwerpen van de prachtige kaft van dit proefschrift.

Tot slot wil ik nog een belangrijk persoon in dit hele verhaal bedanken, zo niet de belangrijkste persoon in mijn leven. Mijn steun, mijn toeverlaat, mijn rots in de branding: bedankt, bollie, om al die jaren niet van mijn zijde te wijken en dit hele proces samen met mij te doorlopen. Je hebt me ondersteund als ik dreigde te vallen; je zette me weer met mijn voeten op de grond als ik ging zweven. De afgelopen vier jaren hebben bovendien niet enkel twee doctoraten opgeleverd (eentje van jou en eentje van mij), maar er is ook een huwelijk uit voortgekomen, een huis mee gebouwd en een klein wondertje mee geschied. Er bestaan geen woorden genoeg om te kunnen uitdrukken hoe dankbaar ik voor dat alles ben.

Een doctoraat schrijven was voor mij meer dan alleen mijn vaardigheden als onderzoeker ontplooien. Het was een groeiproces waar jullie mij allemaal in hebben bijgestaan. Nog eens laatste keer **bedankt**, iedereen, om mij te laten uitgroeien van dat kleine welpje tot de collega, vriendin, persoon die ik vandaag ben.

List of Abbreviations

AVT   Audiovisual Translation

BSD   Belgian Standard Dutch

CBD   Colloquial Belgian Dutch

DPC   Dutch Parallel Corpus

RBBN   Referentiebestand Belgisch Nederlands

VRT   Vlaamse Radio- en Televisieomroeporganisatie

List of Tables

List of Figures

Table of Contents

Chapter 1
Introduction

Over the past few decades, differences between translated and non-translated language have been a prevalent topic within the field of Corpus-based Translation Studies (e.g. Baker 2004; Dayrell 2007; Malmkjaer 1997; Mauranen & Kujamäki 2004). More specifically, it has been investigated to what extent translations and non-translations differ on a lexical (e.g. Laviosa 1998), grammatical (e.g. Olohan & Baker 2000) or discursive (e.g. Hatim & Munday 2004) level, and how these differences can be explained. Traditionally, it is believed that there is a general tendency to normalize or standardize translations, which is for instance illustrated by the fact that more neutral expressions, more conventional and less creative language is used in translations, compared to their source texts or comparable non-translated texts. However, there has recently been a clear focus shift from the superficial charting of linguistic differences between translations (as a whole) and non-translations (as a whole) to a more fine-grained, register-sensitive and source-language-sensitive approach (e.g. Bernardini & Ferraresi 2011; De Sutter et al. 2012a; 2017; Kruger & van Rooy 2012). These studies have clearly shown that this standardization tendency is just a tendency, and not a universal, as the specificity of language use in translations does not only depend on their translational status, but also on a number of other factors such as genre, source language and the target audience. De Sutter et al. (2012a) and Kruger & van Rooy (2012) have demonstrated this, for instance, for the use of formal lexemes and structures in Dutch and South African English respectively. Furthermore, Evert & Neumann (2017) have demonstrated that less-prestigious languages (e.g. German) are more tolerant toward source language interference (or shining through) than more prestigious languages (e.g. English), since the translations into German seemed to accommodate more characteristics of English as the source language than translations in the opposite direction. Recent empirical studies also attribute great importance to the impact of editorial intervention on translation strategies (Kruger 2017) and to the effect of computer-aided translation (Čulo et al. 2017). Kruger (2017), on the one hand, showed that edited translations are not only more explicit

(e.g. they contain more non-contracted word forms, complements and linking adverbs) than non-edited translations, editors also prefer conventionalized and standardized language use in translated texts. On the other hand, Čulo et al. (2017) suggested that post-edited translations are largely influenced by the initial output of translation software, as the machine translation product often seems to "shine through" in the post-edited texts: "the P[ost-] E[dited] products tend to exhibit variation where the M[achine] T[ranslation] products exhibit variation, they show similar peaks of perplexity values [i.e. normalized language] for the same term, and there is no variation when there is none in MT" (my addition; Čulo et al. 2017: 197).

On an explanatory level, the aforementioned research has greatly contributed to a better understanding of the sociolinguistic motivations underlying the translators' behavior, more particularly in terms of risk aversion, in which *risk* is interpreted as "the probability of an undesired outcome" (Pym 2005: 34). In this context, Pym (2005) describes translators as mediators in communication that want to reduce the risk of not successfully achieving their communicative goals. As a consequence, they are held to be more risk-averse than other professional language users, since "translators will tend to avoid risk by standardizing language and/or channeling interference, if and when there are no rewards for them to do otherwise" (Pym 2008: 326). In other words, translators will prefer to use a safe variant (e.g. a variant that is widely accepted as a standard variant), instead of a risky variant (e.g. a variant that is considered an informal non-standard variant). In a recent study based on a corpus of English translated from Afrikaans, Kruger (in press) demonstrates that translators tend to use the explicit complementizer *that*, even though the African language prefers to leave *that* implicit. Since the use of this explicit complementizer is assumed to be strongly correlated with formality (Biber et al. 1999) and is often retained to guarantee comprehensibility, particularly in complex discourse (Torres Cacoullos and Walker 2009), this translation strategy is seen as "a collateral effect of a conservative preference for a more formal style, motivated by risk avoidance" (Kruger in press).

With regard to the Dutch language area, Delaere (2015) has demonstrated that translators of 'regular' written genres more often opt for commonly accepted Standard Dutch words and constructions in comparison to writers of original texts (non-translators), which exemplifies a standardizing, norm-adhering trend, resulting from the translators' risk-averse behavior (cf. also Delaere et al. 2012; De Sutter et al. 2012a, and Delaere & De Sutter 2013). Simultaneously, it was found that translators' behavior is not uniform at all, as norm-adherent lexical and grammatical choices appeared to significantly depend on extralinguistic factors such as source language, target audience and register or genre: "registers with a lot of editorial control (fiction, non-fiction and journalistic texts) contain more standard language than the less edited registers (administrative texts and external communication)" (Delaere et al. 2012: 203). However, the aforementioned research is based on a limited number of translated and non-

translated texts, which prevents us from generalizing the results across other text types. Subtitles, for example, are characterized by a colloquial input and are thus situated between oral and written language. Unlike the text material of Delaere (2012)[1], subtitles are the result of the transformation of spoken language, with its typical colloquial and often non-standard features, into written text. This might encourage subtitlers to use non-standard linguistic items (Díaz-Cintas 2010: 344-346; Karamitroglou 2000; Neves 2004). Nevertheless, subtitles are also heavily edited translations, which might stimulate the use of standard language. As a consequence, the question arises what kind of linguistic choices are made by subtitlers.

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; Remael et al. 2008; Rosa 2001). Although these studies offer a number of valuable first insights into language variation in subtitling, they have a small empirical base (the subtitled material was collected from, respectively, two films, one film and four episodes selected from four different series). Moreover, the aforementioned studies all adopt a qualitative approach, without applying statistical techniques to analyze linguistic variation in the subtitles, which makes it difficult to find patterns in the subtitlers' linguistic choices. Furthermore, no attention went to lexical, morphological, syntactic, and pragmatical variation within the subtitles, nor was the influence of contextual factors, such as genre, on the language used in subtitling taken into consideration. Nevertheless, it is crucial to acknowledge the potential influence of these factors, as previous studies on language variation have already demonstrated their relevance (e.g. Delaere et al. 2012). 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. It was 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 only focused on Colloquial Belgian Dutch lexicon (thereby omitting colloquial syntactic and morphological constructions), and it did not investigate

---

[1] Some texts in the corpus of Delaere (2012) have also a colloquial character (e.g. interviews, political speeches). Nevertheless, the number of these text types is limited.

the Flemish spoken source text and its potential influence on the subtitlers' linguistic choices. Furthermore, there was only one television genre taken into consideration, which makes it difficult to verify which contextual factors determine the subtitlers' linguistic choices. This dissertation therefore sets out to quantitatively investigate linguistic variation in subtitling, using a large corpus of interlingual and intralingual subtitles that allows us to examine the subtitlers' linguistic choices in various contexts. In addition, these findings are compared to the linguistic choices made in other written translations and non-translations. In order to understand the linguistic choices Flemish subtitlers make, a qualitative analysis will be performed by means of semi-structured interviews, evaluation reports, and observational data that were collected at the subtitling department of the Flemish public broadcaster.

The specific language situation in the (officially) Dutch-speaking part of Belgium, also called Flanders, makes this research particularly interesting. In Flanders, Belgian Standard Dutch (*Belgisch Standaardnederlands*) is the official variety that is generally accepted and especially used in very formal spoken registers and in written language. In less formal and informal contexts, however, language use in Flanders is characterized by typical lexical and grammatical features that are widely used, but not accepted as Belgian Standard Dutch by the language authorities (e.g. Dutch Language Union[2], Hendrickx 1998). This non-standard informal variety is known as *tussentaal* (literally: 'in-between language'), and less often also termed Colloquial Belgian Dutch (e.g. Geeraerts & Van de Velde 2013; Ghyselen & Van Keymeulen 2016), a term that we will also use in this dissertation. 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. The Flemish language situation has attracted considerable attention, especially from the 1980s onwards, with many synchronic and diachronic usage-based studies focusing on linguistic differences and similarities between Belgian Standard Dutch and Netherlandic Standard Dutch (i.e. the Standard Dutch variety used in the Netherlands) on the one hand, and on the relationship between Colloquial Belgian Dutch and Belgian Standard Dutch on the other hand (e.g. De Caluwe 2002; 2009; Geeraerts et al. 1999; Impe 2006; van de Velde 1996). Other Dutch linguists have also focused on the context in which these different language varieties are used and on their (sociolinguistic) function in daily usage (e.g. Absilis et al. 2012; Jaspers 2001; Van Hoof & Jaspers 2012). Underlying motivations of this research trend can be found in the

---

[2] See http://taaladvies.net

particular standardization process of Dutch in Belgium in the late 19ᵗʰ century and its explicit normative orientation toward the Standard Dutch variety as used in the Netherlands (Taeldeman 1992). At that moment, language policy makers wanted to clear the language of the typical Belgian-Dutch variants (that were considered dialectal or regiolectal) and have them replaced by the Netherlandic-Dutch variants (which were assigned the status of standard language) in order to create a common, supraregional standard language, viz. General Standard Dutch. These attempts were partially successful in the more formal registers which have converged largely toward the northern variety (Geeraerts et al. 1999). However, this Dutch standard language is not completely uniform in both parts of the language area, since there are some quite noticeable pronunciation differences in addition to lexical and grammatical differences. Recently, the language policy in Flanders has become more tolerant toward the Belgian variety of Dutch and increasingly considers Belgian Standard Dutch a variety to be used in formal contexts and in written language (De Caluwe & Van Renterghem 2011). In the informal registers, the remaining typical lexical and grammatical features of Colloquial Belgian Dutch are nowadays tolerated to some extent, but they are not fully accepted. Consequently, professional writers (including translators and subtitlers) keep struggling with the status of these Colloquial Belgian Dutch variants (e.g. Delaere et al. 2012; Remael et al. 2008). This linguistic tension forces them to continuously evaluate the status of words, constructions and idioms that include colloquial features that are frequently used, although they are not accepted by the official language authorities. Furthermore, research dating from the beginning of the 21th century showed that the considerable distance between (Colloquial) Netherlandic Dutch and Belgian Dutch as well as between informal and formal Belgian Dutch was reflected by the fact that Flemish and Netherlandic fiction programs were intralingually subtitled for viewers of the other part of the language area. Furthermore, dialectal or regional speakers in Flanders were increasingly subtitled in non-fiction programs intended for a general Flemish audience (Remael et al. 2008; Vandekerckhove et al. 2006; 2007). Using a corpus of 793 television programs, broadcast in 2005 by the Flemish public broadcasting company VRT and the commercial channel VTM, Vandekerckhove et al. (2006; 2007) showed that the western regiolect was subtitled more often than the dominant Brabant regiolect on Flemish television. Netherlandic Dutch was almost systematically subtitled.

The particular Flemish linguistic situation has driven Dutch-speaking public media to develop a language policy, with specific guidelines for television and radio hosts. The public broadcaster *VRT*[3], for instance, strongly clings to the use of standard language in its policy, aiming to be 'the norm for the Belgian variety of the Dutch standard language'

---

[3] *VRT* stands for *Vlaamse Radio- en Televisieomroeporganisatie*, the Flemish public broadcaster for radio and television in Flanders. Website: www.vrt.be

and 'therefore adopting an attractive, clear and correct standard language that takes into account and is adjusted to its audience' [my translation] (Hendrickx 1998: 1; see also Hendrickx 2007). According to these guidelines, the public broadcaster only occasionally allows the use of spoken non-standard, regionally colored varieties, such as Colloquial Belgian Dutch, in order to preserve the 'credibility of the characters' [my translation] (Hendrickx 1998: 4); although the default language variety to be used, especially in informative programs, such as news and documentaries, is Belgian Standard Dutch (Hendrickx 2012). Nevertheless, recent studies have shown that non-standard varieties (e.g. Colloquial Belgian Dutch, regiolect, dialect) are frequently spoken on Flemish television, even in programs that would actually require Belgian Standard Dutch (e.g. Prieels 2013; see also the evaluation reports of the project *Taalhantering* of KU Leuven & VRT 1999). In television fiction, the use of colloquial language is even highly stimulated, not only by the program makers and the actors, but also by the policy makers, to retain the authenticity and the credibility of the program and its characters (Van Hoof 2010; 2015). The aforementioned studies suggest that the established norm regarding the language used on television has changed. These language dynamics are often attributed to a process of destandardization, which involves that 'the established standard language loses its position as the one and only "best language"' (Coupland Kristiansen 2011: 28; see also Grondelaers & Van Hout 2011; Grondelaers et al. 2011 for the Dutch language area). Nevertheless, the influence of these spoken language changes on the written language has barely been taken into consideration. Given the recent linguistic dynamics in the spoken standards, however, this raises the question whether these processes also affect written language. To fill this gap, the present study will offer an insight into the (de)standardization tendencies in written subtitles on Flemish television. More specifically, we will investigate which variety is used by subtitlers, considering that subtitles are written reproductions of spoken language with its typical colloquial features. Furthermore, VRT's language policy documents date from almost ten years back, which makes it particularly interesting to investigate to what extent the actual subtitling practice conforms to the prevailing norm.

Research goals

The main goal of this dissertation is to investigate how subtitlers deal with the norm-related linguistic tension that exists between VRT's language policy on the one hand, and the specific Flemish language reality on the other hand. More concretely, this study will investigate to what extent subtitlers in Flanders choose Colloquial Belgian Dutch words and constructions instead of Belgian Standard Dutch words and constructions, whether

their choices differ in interlingual and intralingual subtitles, and how these choices can be compared to the linguistic choices that are made in Dutch translations and original Dutch texts. In addition, we will verify which contextual factors influence the use of Colloquial Belgian Dutch variants vs. Belgian Standard Dutch variants by applying multivariate statistical techniques. Contextual factors that will be included in the research design are *source language* (English vs. Belgian Dutch vs. Netherlandic Dutch), *program genre* (e.g. fiction vs. documentaries), *program purpose* (e.g. informing vs. entertaining), *target audience* (adults vs. children), and *cast* (e.g. actors vs. non-actors). It can be assumed that in certain contexts, more Belgian-Dutch features will show up, because the need to conform to the Standard Dutch norm can be considered less outspoken. Moreover, it can be expected that, in certain contexts, subtitlers are more often exposed to Belgian-Dutch colloquialisms, and hence, more frequently triggered to re-use these variants (a case of interference), considering the fact that subtitles are still situated between spoken and written language. By setting up a multifactorial study, and more particularly a profile-based correspondence analysis (cf. Section 3.4), we will investigate how these parameters are related to each other. Furthermore, we will compare the original footage of the TV program to the corresponding intralingual subtitles in order to investigate the extent to which Flemish subtitlers reproduce the Colloquial Belgian Dutch variants. Finally, we will carry out a qualitative analysis of the obtained results, based on semi-structured interviews, evaluation reports, and observational data that were collected at VRT's subtitling department. In doing so, this research will provide insight into subtitlers' attitudes toward VRT's language policy and the language reality in Flanders, which is crucial to understand their linguistic choices in the subtitles. Pursuing the aforementioned objectives will provide answers to a number of specific research questions that are formulated below.

1. **Do Belgian-Dutch subtitles contain more or fewer linguistic features typical for Belgian Standard Dutch than for Colloquial Belgian Dutch in comparison with other translated and non-translated written genres?** Previous research has demonstrated that, in general, translators of 'regular' written genres more often opt for Belgian Standard Dutch words and constructions compared to writers of original texts or non-translations (Delaere et al. 2012; cf. also De Sutter et al. 2012a and Delaere and De Sutter 2013). Based on these results, our study investigates which position Flemish subtitlers take and whether Belgian-Dutch subtitling contains more or less standard language compared to other translations and non-translations.

2. **To what extent do Flemish subtitlers reproduce the spoken Belgian-Dutch colloquialisms in the subtitles or do they even add colloquialisms to the subtitles?** In a recent study, De Ridder (2015) analyzed the use of Belgian-Dutch lexicon 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). These results suggest that Flemish subtitlers (consciously or unconsciously) reproduce the colloquial variants of the original footage of the television program. Nonetheless, De Ridder (2015) did not examine the potential influence of the spoken source text on subtitlers' specific linguistic choices. The present dissertation fills this gap by analyzing the original footage of the TV program to investigate how often the subtitlers reproduce or translate the spoken Belgian-Dutch colloquialisms. In doing so, we will be able to verify the effect of the spoken source text on the language that is used in the subtitles in order to substantiate the aforementioned assumptions.

3. **To what extent do Flemish subtitlers adhere to the language policy of the Flemish public broadcaster, regarding the use of colloquial lexemes and colloquial grammatical constructions?** VRT's subtitling guidelines explicitly tolerate the use of colloquial lexicon in fiction programs (e.g. soaps), whereas colloquial grammatical constructions must be converted into standard language. This raises the question as to whether this means that subtitlers exclusively reproduce colloquial lexemes, whereas colloquial grammatical constructions are converted into standard language. Previous research focused predominantly on lexical features of standard and non-standard language (e.g. De Ridder 2015), whereas little or no attention went to the use of colloquial syntactic and morphological constructions in the subtitles. This dissertation will therefore verify whether the subtitlers also reproduce grammatical colloquialisms. Moreover, case study 3 investigates the exact proportions of lexical vs. grammatical standard and non-standard (colloquial) features in the subtitles.

4. **Which contextual factors (e.g.** *source language, program genre, target audience*) **have an influence on the subtitlers' linguistic choices?** Delaere et al. (2012) found that translators' behavior is not uniform at all, as their lexical and grammatical choices significantly depend on the source language, the target audience and the text genre. This raises the question which contextual parameters (*program genre, purpose, target audience, cast*) affect the subtitlers' norm-related linguistic choices.

5. **How do the subtitlers themselves explain their linguistic choices?** To understand the language used in subtitling, it is necessary to get an insight into the subtitlers' ideas and opinions regarding 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.

Structure of the thesis

This dissertation is divided in two main parts. The first part covers the theoretical background (Chapter 2) of this research; the second part consists of three empirical case studies (Chapters 3 to 6).

Chapter 2 provides the theoretical and empirical background information against which this research is situated. In Section 2.1, the medium *audiovisual translation* will be discussed by giving an overview of the related literature as well as by focusing on the contribution of this study to existing research. After clarifying the concept of a *norm*, Section 2.2 will elaborate on some norm-related key concepts, such as *standard language*, *(standard) language ideology*, *(de)standardization* and *demotization*. Section 2.3 outlines the Dutch standardization process, together with its impact on the current language situation in Flanders and the language policy of the Flemish public broadcaster.

Chapter 3 provides some information concerning the methodology that was applied in the study of VRT television subtitles. First, some general hypotheses are formulated and the different corpora are presented. Next, the selection process of the linguistic variables and annotation procedures are discussed. Finally, this section ends with a profound description of the statistical techniques and the qualitative approach that were used.

Chapters 4, 5 and 6 present the three case studies that were carried out in this dissertation. Case study 1 (Chapter 4) measures the extent to which the language used in subtitles, in comparison to regular written translations and non-translations, conforms to explicit linguistic norms. In case study 2 (Chapter 5), we will investigate whether the subtitles contain more colloquial lexemes than colloquial grammatical constructions, and which contextual parameters influence these linguistic choices. Finally, case study 3 (Chapter 6) examines to what extent Flemish subtitlers reproduce Belgian-Dutch colloquialisms from the spoken source text in the subtitles and how they explain these linguistic choices.

Finally, Chapter 7 summarizes the overall results and provides a general discussion and conclusion. Furthermore, some directions for further research are pointed out and an overview of the implications and shortcomings of this study is provided.

Chapter 2
Theoretical and empirical background

The present chapter zooms in on the theoretical and empirical background against which this research is situated. Section 2.1 provides an overview of the existing literature within the field of audiovisual translation (AVT), focusing on various forms of AVT including subtitling, which is the research object of this study. In Section 2.2, we will define a *norm* and we will also concentrate on some norm-related concepts such as *standard language*, *(standard) language ideology*, *(de)standardization* and *demotization*, which are crucial concepts in the study of norm-related language variation in subtitling. Finally, Section 2.3 builds on these general theoretical insights by discussing the prevailing language norms in Flanders. Furthermore, this final section outlines the Dutch standardization process, together with its impact on the current language situation in Flanders and the language policy of the Flemish public broadcaster.

2.1   Existing literature in audiovisual translation

Together with dubbing, subtitling is the most common form of AVT. Traditionally, choosing one or the other form often depended on various economic, ideological and pragmatic motivations. As a result, the AVT landscape was largely characterized by a distinction between so-called 'subtitling countries' and 'dubbing countries'. Nations with a less-used language or a minority language (Dutch, Danish, Greek, etc.) generally preferred subtitling, whereas countries with an international language (English, French, Spanish etc.) and a bigger audience favored dubbing practice more often (Gambier 2012). Furthermore, both dubbing and subtitling countries are convinced that their 'own' method is the best one (Bruls and Kerkman 1989; Kilborn 1993; Luyken et al. 1991; Spinhof & Peeters 1999). The most common argument in favor of dubbing foreign TV programs is

that dubbed programs are easy to follow because the audience does not have to read while watching television (Koolstra et al. 2002). Subtitling, on the other hand, is defended with the argument that the original voices of the actors are left intact when the original soundtrack is retained (Mera 1999).

Nowadays, subtitling is in rapid expansion, even in dubbing territories, due to changes in broadcast technology and market pressure for simultaneous release of audiovisual products across multiple territories (Chiaro 2009). As a result, the subtitling practice has recently become a prominent academic field of research. According to Díaz-Cintas (2013: 274), subtitling can be defined as:

> a translation practice that consists of rendering in writing, usually at the bottom of the screen, the translation into a target language of the original dialogue exchanges uttered by different speakers, as well as all other verbal information that appears written on screen (letters, banners, inserts) or is transmitted aurally in the soundtrack (song lyrics, voices off).

In other words, the central concern of subtitling is to render different types of verbal speech information in two lines of concise and intelligible writing with a minimal loss of informative content (Remael 2003). The fact that viewers have to read the written subtitles at a given speed while watching the images at the same time, makes subtitling a challenging practice. As a consequence, subtitles are generally subject to norms of exposure times, reading speed constraints and subtitle density (Gambier 2012). Traditionally, subtitlers were bound by the so-called six-seconds rule (Gielen & d'Ydewalle 1989). Six seconds was the recommended maximum exposure time to keep two lines of each about thirty characters on the screen to let the viewer read the subtitles and to avoid an unnecessary second reading. In a recent study, however, Szarkowska (2016) has taken down this six-seconds rule, since she has demonstrated that the reading speed of the viewers is actually much higher (viz. up to 84 characters in six seconds). As a consequence, many broadcasting stations and online streaming services (e.g. Netflix) use subtitles of more than 64 characters. Furthermore, the viewer is able to compare the subtitles to the spoken source language, as the subtitles appear simultaneously with the spoken television dialogue. This 'feedback effect' makes subtitles a 'vulnerable' text type (Díaz-Cintas & Remael 2007: 55), as the subtitlers' work runs the risk of being evaluated negatively by the public if the subtitles deviate too much from the spoken dialogue.

Within the field of subtitling, a distinction is traditionally made between interlingual and intralingual subtitles. Interlingual subtitles, on the one hand, are used to translate a foreign dialogue into subtitles in the native language of the audience. These subtitles are usually open subtitles, i.e. they appear automatically on the television screen. Intralingual subtitling, on the other hand, is done in the same language as the original dialogue of the audiovisual product, and therefore merely involves 'a shift from the spoken mode of the verbal exchange in a film or TV programme to the written mode of

the subtitles' (Gambier 2012: 49). As a result, intralingual subtitles are sometimes also called *same language subtitles* (SLS), as they are (free) transcriptions of the spoken source language. Although intralingual subtitling can also appear automatically on the screen, it is often a teletext option on television. Such optional subtitles are called closed subtitles. Especially when the audience is hearing-impaired, intralingual subtitles can support these viewers while watching television. Such intralingual subtitles for the deaf and hard-of-hearing respect a certain degree of synchronization by reproducing to a large extent the lexicon and syntax of the original speech, because many hard-of-hearing people lip-read as an additional source of information. In addition, subtitles for the deaf and hard-of-hearing also render non-verbal audio material (e.g. whistling, knocking on a door) into text (Gambier 2012).

In brief, AVT may be considered a complex and diversified domain (Rosa 2016). Not only does the combination of audio, visual and verbal signs make the audiovisual text itself a very complex multimodal product, all of these components in the audiovisual source text have to be translated to the audiovisual target text too (Gambier 2012; Zabalbeascoa 2008). A considerable amount of research within the field of AVT went to the exploration of strategies that are used for coping with the information load in the original text. For dubbing, for example, Zubiria (2012) demonstrated that reduction and modulation are frequently used techniques to synchronize the source and target text, whereas Remael (2007) showed that subtitles are frequently abbreviated in order to deal with the original speech tempo. On the other hand, Szarkowska (2005) found that subtitles are more explicit than the source text, for instance by inserting vocatives as a means to distinguish between different characters. Other studies in AVT have explored specific linguistic characteristics in audiovisual translation. Baños (2013), for instance, found that the adverbial intensifiers *very*, *so*, *totally*, *pretty*, *really* in English-to-Spanish dubbed sitcoms are most frequently translated by means of degree adverbs (e.g. *muy/tan* 'very' and *mucho/tanto* 'many'), making the dubbed speech less speech-like and more similar to written language. In her study on phrasal verbs in original Italian films and their French-to-Italian dubbed versions, Valentini (2013) argued that dubbed language is both lexically and grammatically poorer than original language, as the dubbed language contains fewer verb-particle constructions. Some recent investigations have also tackled linguistic variability, viz. the rendering of geographically colored language varieties, in AVT. In the dubbing of English dialect into Italian, Ranzato (2010) distinguished several strategies (e.g. rhyme, colloquial expressions and non-words) to translate Cockney's rhyming slang into 'an unlocalised, foreign variant of the standard language', without using an Italian dialect (Ranzato 2010: 121). With regard to the subtitling practice, Ramos Pinto (2017) puts forward a multimodal perspective by focusing on the issue of the translation of non-standard varieties in particular. Furthermore, De Meo (2012) and Tortoriello (2012) both investigated the strategies used in translating Italian dialect into English subtitles. These case studies demonstrated two different trends. Either the

marked dialect intonation was translated by using non-standard grammatical constructions, loan words, idiomatic expressions, etc. (De Meo 2012), or the English subtitles were consistently written in standard English and did not make any attempt at conveying regional varieties (Tortoriello 2012), a technique which can be considered to boil down to standardization. This standardizing tendency is often attributed to the language policy of the subtitling authorities, which generally support the use of a neutral standard language (e.g. Hamaida 2007; Remael et al. 2008; Rosa 2001).

The present dissertation investigates whether this standardizing tendency also counts for the subtitling practice on Flemish television. More specifically, it will be verified whether Flemish subtitlers conform to the language policy of the Flemish public broadcaster VRT, which is largely oriented toward the use of standard language. By examining the extent to which Flemish subtitlers use Colloquial Belgian Dutch rather than Belgian Standard Dutch in various contexts, we will not only contribute to a better understanding of the sociolinguistic motivations underlying the subtitlers' norm-related linguistic choices, but we will also get a better insight into the current position of several standard and non-standard language varieties used in Flanders. In order to get acquainted with the contemporary linguistic situation in the Flemish language area in general as well as on Flemish television, Section 2.3 offers a background sketch of both the historical development and the current position of the Dutch standard language in Flanders. Section 2.2 first defines some norm-related concepts in order to become familiar with the terminology that will be used further on in this dissertation.

2.2   Discussing some norm-related concepts

If we want to set up a study on norm conformity in Belgian-Dutch subtitling, it is indispensable to elaborate on what a *norm* exactly is. In everyday life, norms are assigned great authority because of their regulating function. For instance, if a customer buys a smartphone at the electronics store, he expects that product to meet certain criteria (viz. it must be able to send text messages, to make phone calls and to surf the Internet). If the product does not match these expectations, the buyer will feel deceived. To avoid this kind of disappointment, statutory regulations are introduced to protect the consumer.

Beside these technical norms, however, it is important to emphasize the (even more essential) existence of social norms, which largely regulate human behavior in society. In daily interactions, people have mutual expectations, as they do not only expect a certain behavior of the others, but they will also behave in a way they assume others will expect them to do. In this context, Gloy (1975: 40) talks about 'Erwartenserwartungen' or 'expectations of expectation' (Bartsch 1984b: 368; see also Giddens 1984; Grice 1969;

Luhmann 1972; Schiffer 1972; Ullman-Margalit 2015). A good example are norms of communication. In order to establish successful communication, it is important that 'one is able to rely on the other(s) acting according to the valid norms' (Bartsch 1984b: 368; see also Jaspers 2010). For example, if person A assures that he will pay a visit to person B tomorrow, person B expects person A to keep his promise and to pass by the next day. These norms offer us an insight into the behavior of others and serve as guidelines for our own actions in social interactions. As a result, people will largely stick to these norms, because they provide a certain sense of security. This could explain why people often react emotionally when their norms are at issue (Bartsch 1984a; for an extensive theoretical discussion on linguistic norms, we refer to Bartsch 1987). Deviant behavior is considered as threatening and often causes social conflicts. For instance, if person B in the previous example does not show up the other day, person A will feel disappointed. As a result, person A gets angry with person B, a consequence that follows as person B's behavior does not conform to the social norm, i.e. he did not keep his promise. As Bartsch (1984b: 368) states: 'People care about keeping to the patterns, and they correct deviancies'. According to Bartsch (1984a), criticism, corrections, and sanctions are aspects that establish the normative power of a norm. In this respect, norms are similar to prescriptions or commands. However, a subject follows the prescriptions or commands of an authority, whereas the norm subject identifies itself with the norm authority and vice versa. In other words, a norm applies to everyone, both the subject and the authority, whereas a prescription does not necessarily apply to the authority itself.

Social norms generally hold an implicit power, which means that people unconsciously act in accordance with the norm, even though there are no official prescriptions. Nevertheless, sometimes it may be useful for the controlling authorities to formulate an explicit codification to define the content and the scope of the norm (Marklund Sharapova 2000), especially with regard to linguistic norms. In the context of language planning, for instance, standard languages require an explicit codification to impose a certain uniformity regarding correct pronunciation, lexicon and grammar. These specific language norms are generally written down in dictionaries and grammar books. Especially in communities in which the standard language is the result of the adoption of an exoglossic variety (cf. Grondelaers et al. 2016 for the Dutch-speaking part of Belgium), strict codification is needed to guarantee the implementation of the imposed norm. According to Milroy (2001: 530), most European countries have 'standard language cultures', which means that in these speech communities, language exists in a standardized form. Underlying motivations for this standardization process can generally be found in societal interests and beliefs such as increasing mobility, scientific or economic advancement, and religious or political ideologies. For instance, the need for a uniform, correct and prestigious variety often comes with the desire to assert a nation's cultural identity, which is typical of a nationalistic ideology (cf. Haarman 1997; see also Bauman & Briggs 2003; Deumert & Vandenbussche 2003a; 2003b; Kroskrity 2000). This

standardization process is driven by the idea that there is a 'best' language (Coupland & Kristiansen 2011: 28). As a result, standard language is often defined in evaluative terms, such as the 'best variety' (Grondelaers et al. 2016: 120), 'proper ways of speaking' (Mugglestone 2003), and 'refined talk' (Coupland & Kristiansen 2011: 12; see also Jaspers 2001 and Jaspers & Brisard 2006 for an analysis of the linguistic discourse on standard language). Furthermore, language users generally consider standard varieties to be more 'beautiful' than non-standard varieties (see for instance Bishop et al. 2005; Coupland & Bishop 2007; Giles 1970; Grondelaers et al. 2011; Trudgill & Giles 1978; Van Bezooijen 2002;). Linguists, however, generally characterize *standard language* as 'a slippery concept' (Coupland and Kristiansen 2011: 11), which is 'difficult to delineate and define with any amount of precision' (Grondelaers et al. 2016: 120). As a result, it is challenging to indicate in what a standard language distinguishes itself from other (non-standard) language varieties. Traditionally, the definition of a standard language is based on Auer's (2011: 486) three criteria:

> (a) a standard language is a common language, i.e. one which (ideally) shows no geographical variation in the territory in which it is used; (b) a standard variety is an H variety, i.e. it has overt prestige and is used in situations which require a formal way of speaking [...] and writing; and (c), a standard variety is codified, i.e. 'right' or 'wrong' plays an important role in the way in which speakers orient towards it.

First, standard languages are assumed to be uniform and invariable, which means that they are fixed by a set of highly prescriptive rules 'from which any deviation is forbidden' (Smith 1996: 65-66). Second, the standard language variety is often considered the highest prestige variety. Especially during the 18th and 19th centuries, standard languages developed into symbols of 'educatedness' and 'refinement' (Deumert & Vandenbussche 2003b: 459). They were seen as the language varieties that are used by the highest social layers of a population. In this respect, the higher social classes often played a key role in the standardization process of a language, since the promoted standard variety was generally based on the language used by this elite. In addition, Milroy (2001) indicates that language varieties do not actually have prestige in themselves, but they acquire prestige from the perceived high status of its speakers. Finally, the standard language variety is regarded as the correct variety. This belief holds that, 'when there are two or more variants of some word or construction, only one of them can be right' (Milroy 2001: 535). In Dutch for example, many language users drop the *te* element in the construction *beginnen + te + infinitive* 'to start to'. However, the construction with *te* is the only correct form. If someone uses the variant without the *te* element, his language use does not conform to the Standard Dutch norm.

One of the main disadvantages of Auer's (2011) criteria is that standard languages are thus represented as 'non-vital, virtual' varieties that are merely a point of reference for

correct language use (Grondelaers et al. 2016: 128). In everyday speech, however, this standard ideal is inevitably affected by linguistic variability and, in this respect, never produced in its purest form. In addition, Smakman (2006) has found that even the speech of newsreaders, who are considered the 'best' speakers of Dutch by a large group of informants, shows extensive phonetic variation, which demonstrates that standard languages are 'inherently variable and can never be fully standardized' (Milroy & Milroy 1985: 22). In other words, standard languages are 'an idea in the mind rather than a reality, a variety that is never perfectly and consistently realized in spoken use' (Milroy 2001: 543). In this respect, standard languages are, as it were, ideological idealisations. The term 'standard language ideology' (SLI) was coined by Milroy & Milroy (1985: 23) to refer to 'a set of abstract norms to which actual usage may conform to a greater or lesser extent'. Building on the anthropological theories about language ideologies (e.g. Gal & Irvine 1995; Kroskrity 2000; Silverstein 1979; Woolard 1992), James Milroy and Lesley Milroy are among those linguists who have pioneered ideology-sensitive theories of linguistic standardisation, and their work has provided an important foundation for several recent sociolinguistic studies on the nature of language standardization. With regard to Europe, the establishment of the LANCHART (Language Change in Real Time) Centre at the University of Copenhagen has broken new research ground. A central initiative of this centre was the implementation of the SLICE (Standard Language Ideology in Contemporary Europe) programme, 'a European network of like-minded researchers, with the prospect of developing one or more innovative, comparative European projects on standard languages, linguistic standardization and linguistic destandardization' (Coupland & Kristiansen 2011: 11). With this project, the SLICE network aimed to describe the ongoing standard language dynamics throughout Europe.

As a consequence of growing immigration, globalization, democratization and informalization, the uniformity – and even the standardness – of standard languages all across Europe has recently come under pressure (Odendaal 2014; Ghyselen et al. 2016). These societal trends have led to more variability within languages and language varieties, which indicates that the 'belief in the blessings of linguistic unity may not be so strong anymore' (Coupland & Kristiansen 2011: 29). The studies that were conducted by the researchers of the SLICE programme focused on two hypothetical linguistic scenarios, suggested by Coupland and Kristiansen (2011): *destandardization* and *demotization*. The term *destandardization*, on the one hand, is used to refer to 'a possible development whereby the established standard language loses its position as the one and only "best language"' (ibid: 28). This process can affect 'value levelling' between different varieties, which can lead to 'a radical weakening, and eventual abandonment, of the "standard ideology" itself' and to a greater tolerance toward language variation in all kinds of situations (ibid.). In a context of *demotization* (cf. *demotizierung* by Mattheier 1997), the 'increasing variability does not challenge or threaten the standard language ideal, but it stretches the standard to include regional and social variation' (Grondelaers et al. 2011:

200). In other words, the standard language continues to be the 'best' variety, but with an extended interpretation of what this standard language exactly is. Language users still have the intention to speak the standard variety, but this variety incorporates more variability.

Processes of destandardization and demotization have recently been observed in countries all across Europe. Kristiansen (2009, cited in Coupland and Kristiansen 2011: 27) has conceptualized these dynamics as a continuum, with at the one pole, 'countries [...] with strict and strong standard languages'. Iceland, France, and Denmark, for instance, are nations where the standard variety has hold a strong position for a long time. Although the Icelandic standard is kept variation-free until today (Árnason 2003), the standard varieties in France and Denmark are currently under pressure. In France, on the one hand, the French standard language ideology is increasingly challenged due to growing linguistic diversity and multilingualism (Lüdi 2012). On the other hand, Standard Danish nowadays includes features that are generally associated with low-status ('popular') Copenhagen speech. Especially Danish youngsters evaluate this new way of speaking more positively than the traditional high-Copenhagen accent, as well as the locally colored accents of Copenhagen speech (Kristiansen 2003b). In this respect, the belief that there is a 'best language' is not abandoned, but 'the "low-Copenhagen" accent is replacing the "high-Copenhagen" accent as the "best language"' (Kristiansen 2001, cited in Coupland & Kristiansen 2011; see also Kristiansen 2003a; 2003b), thus corresponding to a demotization scenario. At the other pole of Kristiansen's (2009) continuum, we find countries of which 'it is an open issue whether [... they] have a standard language at all' (my addition; Coupland and Kristiansen 2011: 27). For instance, the strong position of the traditional Norwegian dialects in everyday language, but also in more formal contexts, has repeatedly opened the debate on the questionable existence of a standard language in Norway (e.g. Jahr 2003; Hilton 2010; Sandøy 2011). Between these two poles, countries with relatively strong standard languages can be found, although the ideological character of these standards is changing. Next to the supra-regional (written) German standard, for example, two varieties are dominating the German speech community: a regional language (*Regionalsprache*) for informal situations and a regional standard (*Regionalstandard*) for more formal contexts (Mattheier 2003). Processes of demotization have recently caused the loss of the regional variants in favor of more general 'allegro' forms (e.g. the deletion of final-*t* in *ist* ['is'] or *nicht* ['not']), thus making the oral standard more homogeneous. According to Auer & Spiekermann (2011: 174), this could eventually result in the abandonment of the 'old-fashioned' pronunciations of *ist* and *nicht* in favor of the new allegro variety, since many Germans consider this colloquial standard 'the language they grew up with'. In Britain, on the contrary, the ideological power of Standard English is challenged by the increasing use of the non-standard varieties, especially in British media. On some radio channels and in certain genres or formats (e.g. stand-up comedy), it is even believed to be 'laughable' if

the program is hosted by a standard-speaking presenter (Garrett et al. 2011: 63). As a result, this destandardization tendency has induced an increasing use of English vernaculars such as 'Estuary English', an intermediate variety between the standard language and the London-based Cockney dialect (cf. Mugglestone 2003; Rosewarne 1984).

The aforementioned studies have not only demonstrated that standard languages currently seem to be under pressure all across Europe, but they have also shown that the specific character of these linguistic processes differs from nation to nation. Furthermore, the preference for standard or non-standard language varieties has become increasingly context-dependent and the need to obtain general acceptance of only one 'best language' variety seems to be over (see also Coupland 2010; 2014). In this respect, broadcast media are believed to play a central role in carrying forward social and sociolinguistic changes, especially with regard to standard and vernacular spoken varieties (Coupland & Kristiansen 2011; Coupland 2014; 2017). Initially, television and radio were widely considered to be guardians of standard languages. In Britain, for example, the national broadcaster BBC took on the explicit role of promoting the 'best language' for several decades (Bell 2011). In Flanders too, the Flemish public broadcaster VRT adopted an exemplary role in the promotion of Standard Dutch (Hendrickx 1998). But as important as radio and television initially were for the conservation of a strict and well-articulated standard norm, just as important have they nowadays become as models of linguistic diversity and distributors of new words and expressions (Thelander 2011). Several studies have shown, for example, that the language of newsreaders and television hosts increasingly shows traces of non-standard, colloquial variation (e.g. Bell 2011; Hedin 2009; Soukup 2016; Thelander 2011). In this context, AVT has become a particularly interesting research domain. Not only does AVT make it possible to broadcast films and television programs around the world by converting spoken movie dialogues into the native language of its viewers, but it can also translate regionally colored varieties of the source text into a standard target language. Intralingual subtitling, for instance, can help the audience to understand speakers using a regionally limited dialect that may not be intelligible to all viewers (cf. Section 2.1). The question arises, however, whether the aforementioned linguistic dynamics also affect the language that is used in the subtitles. In other words, do subtitles contain traces of these destandardization or demotization processes, considering that they are written reproductions of colloquial language?

As mentioned in the introduction of this dissertation, the use of regionally colored varieties on television has recently become a core subject in linguistic research. A particular case can be found in the Dutch language area. Even though Flanders and the Netherlands officially share the same language (viz. Standard Dutch), both regions have their own standard and non-standard varieties with considerable phonetic, lexical and grammatical differences. As a result, it is common practice for Dutch TV channels to subtitle Flemish TV shows. Likewise, Dutch shows broadcast on Flemish channels are often subtitled there as well. Furthermore, dialectal or regional speakers in Flanders are

often subtitled in TV programs intended for a general Flemish audience (Remael et al. 2008; Vandekerckhove et al. 2006; 2007). In this context, it is particularly interesting to consider Belgian-Dutch subtitling a subject of research, since Flemish subtitlers have to choose whether they reproduce the colloquial variants from the spoken source text rather than converting these colloquialisms into standard language. By investigating the subtitling practice in Flanders, we hope to find out more about the current position of several standard and non-standard varieties in the Dutch language, and whether these dynamics can be attributed to a process of destandardization or demotization. The next section sketches the historical background to the contemporary linguistic situation in Flanders, together with its influence on the language policy of the Flemish public broadcaster.

2.3 Norms and standardization in the Dutch-speaking language area

Within the field of Dutch linguistics, a considerable amount of literature is dedicated to the historical development of Dutch in both the Netherlands and in Flanders. This section, however, will be limited to a concise discussion of the language situation in the Dutch language area[4]. Together the Netherlands and Flanders constitute the main part of the Dutch-speaking language area in the world, which is due to particular historical developments nowadays characterized by a complex language situation. Although Flanders and the Netherlands officially share the same supraregional standard, viz. Standard Dutch, there have always been noticeable pronunciation differences as well as lexical and grammatical diversity between north and south. The consensus among linguists nowadays is that besides a common 'Standard Dutch' core that is shared by Flanders and the Netherlands, there are two area-specific standard languages, viz. Netherlandic Standard Dutch and Belgian Standard Dutch. This recognition of the 'pluricentric' nature of the Dutch language area (De Caluwe 2005: 53), with north and south each constituting a legitimate language centre, is however fairly recent, and was preceded by a period in which linguists and intellectuals in Flanders advocated close convergence to the norm for Standard Dutch as it had been developed in the north.

---

[4] This background information is primarily based on the works of Janssens and Marynissen (2008) and Willemyns & Daniëls (2003).

2.3.1   A brief historical outline

While the Netherlands started to develop a prestige variety shortly after having gained their independence in 1648, Dutch became a standard language in Flanders more than two centuries later. Until the foundation of Belgium in 1830, the southern part of the Low Countries was under the government of Spanish, Austrian and French authorities and was thus politically, culturally, and linguistically separated from the Netherlands. Traditionally, it is believed that, during this period of foreign rule, the southern provinces adopted French for all supraregional communication, whereas the use of Dutch was reduced to the rural dialects. It was merely from the late 19th century onwards, when the so-called Flemish Movement started to combat the suppression of Dutch in Flanders, that the Dutch language was put forward to serve as an official variety for government, culture, and education. However, since the use of Dutch in Flanders was believed to have been restricted to the rural dialects for centuries, it was considered unsuitable for supraregional use. Instead of further developing the Flemish rural dialects into a unified standard language, Flemish intellectuals decided to take over the standard language of the Netherlands in all its registers (Taeldeman 1992:42–44), rendering Netherlandic Dutch the dominant variety in the Dutch language area. It should be noted, though, that more recent theories (e.g. Vosters & Vandenbussche 2008) state that until the French domination, and even under Spanish and Austrian rule, Dutch was used as a (semi-)official language in the South, next to French, and that also the higher social classes used Dutch in their written communication (instead of having been completely frenchified). Furthermore, there were multiple attempts to standardize written Southern Dutch during the 18th century. According to Vosters et al. (2010), the traditional, negative view of the miserable state of the language in Flanders should therefore be seen as a discursive strategy of the integrationists, who aimed for the adoption of the Northern Dutch standard norm in Flanders (see also Rutten and Vosters 2010; Vosters 2013; Van Hoof 2015).

As a result, the language policy in Flanders was explicitly oriented toward the north in the late 19th century. Language policy makers wanted to clear the language of the typical Belgian-Dutch variants (that were considered dialectal or regiolectal) and have them replaced by the Netherlandic-Dutch variants (which were assigned the status of Standard Dutch), in order to create a common standard language. As a consequence, this period of *hyperstandardization* involved an intensively 'propagandistic, large-scale and highly mediatised linguistic standardisation campaign that has thoroughly ideologized and hierarchized language use in all corners of Flemish society' (Van Hoof and Jaspers 2012: 97). Flemings were actively encouraged to adopt the imported Netherlandic standard by means of so-called ABN-campaigns. The use of *Algemeen Beschaafd Nederlands* ('General Civilised Dutch', Willemyns 2013: 143) was not only heavily propagated at school, but also newspapers, radio and television stimulated positive attitudes toward this exoglossic

standard language (Van Hoof and Jaspers 2012; Vandenbussche 2010; De Caluwe 2012a; Van Hoof 2013; 2015). Especially Flanders' public broadcaster VRT (*Vlaamse Radio- en Televisieomroeporganisatie*) adopted an exemplary role in the 'implementation' (Bartsch 1984a: 181) of Standard Dutch by strictly controlling the language that was used by its presenters (Beheydt 1991a; 1991b; Van Hoof 2015; Van Poecke & Van den Bulck 1991). Educative radio and TV programs (e.g. the popular TV program *Hier spreekt men Nederlands* 'Dutch is spoken here', which was broadcast three times a week between 1964 and 1972) were transmitted with the intention of helping Flemish language users to acquire the Netherlandic standard. According to Willemyns (2003: 111) this standardization diligence provided 'almost an entire population in a couple of decades with a more or less new language or, to put it more correctly, with a less known variety of their own language'.

2.3.2   Dutch in Flanders today

In spite of the widespread efforts to propagate Standard Dutch in Flanders, a linguistic gap between the northern and the southern part of the Dutch-speaking language area still exists today. Especially with regard to pronunciation, a different accent clearly distinguishes Belgian Dutch from the Netherlandic-Dutch variety. As a result, the Dutch language area is generally considered as bicentric, as these two varieties of Dutch coexist (e.g. Clyne 1992; De Caluwe 2012b).

In the more formal registers and in written language (in newspapers, magazines, handbooks, official documents, on radio and television, etc.), Belgian Standard Dutch has converged largely (but not completely) toward the northern norm (Geeraerts et al. 1999). As the Belgian-Dutch variety is increasingly recognized as a variety next to the Netherlandic-Dutch variety (De Caluwe & Van Renterghem 2011), the remaining typical lexical and grammatical features of Belgian Dutch are nowadays tolerated to some extent, but they are not fully accepted by the language authorities. As a result, professional writers still struggle with the status of these Belgian-Dutch variants. In informal Belgian Dutch, the language policy has been much less successful, as there is a considerable linguistic distance not only between the Netherlandic and Belgian-Dutch colloquial variety, but also between Belgian Standard Dutch and Colloquial Belgian Dutch (cf. Geeraerts et al. 1999; Goossens 2000; Janssens and Marynissen 2008; Grondelaers & van Hout 2011). Many Dutch linguists attribute the strong position of Colloquial Belgian Dutch to the fierce (hyper)standardization process during the second half of the 20th century, which caused a feeling of language uncertainty among the Flemish people. As a consequence, the Standard Dutch language has never been a familiar medium in which Flemish users felt comfortable. Therefore, Flemings resort to Colloquial Belgian Dutch (or *tussentaal*), when they find themselves in an informal context. The standard variety only seems to be 'appropriate for formal interaction' (De Caluwe 2002: 61) and was given the

function of 'a "Sunday suit", an indispensable piece of clothing which one takes off, however, as soon as the occasion no longer demands it' (Geeraerts 1999 and 2001 cited in Grondelaers et al. 2011; Rys & Taeldeman 2007).

As a result, the contemporary Flemish language situation is characterized by what Auer (2005; 2011) names diaglossia. In Auer's (ibid.) work, the relationship between standard and non-standard language varieties is represented as a cone, in which the tip symbolizes the standard, and the ground circle symbolizes the repertory of non-standard varieties. This relationship is either of a diglossic or a diaglossic nature. 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, Standard Dutch is widely used in the more formal spoken registers and in written language, whereas in the informal contexts and in informal spoken language both standard and non-standard Dutch varieties are used. These colloquial language varieties have been a prevalent topic in numerous heated discussions of Flemish linguists who shared the 'integrationist' vision and had been propagating the use of standard language for decades. Colloquial Belgian Dutch or *tussentaal* was initially considered an inevitable transitional phase, an in-between language, which was 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; Geeraerts 1999; 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 De Caluwe 2006; De Schutter 1998; Goossens 2000; Taeldeman 1992; Van Istendael 1989; Van de Velde 1996). Although the tussentaal debate has stayed alive among linguists 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 21th century onwards, an increasing number of empirical studies started to investigate the use of Colloquial Belgian Dutch in various situational contexts, without passing judgement or condemning this informal variety. For instance, Plevoets (2008) focused on the geographical and situational distribution of some typical morphosyntactic tussentaal features. Lemahieu (2008) studied the language used by young people (aged between 18 and 24 years) in informal conversations. Lybaert (2014a; 2015) set up a perceptual study, in which the informants were asked to evaluate the spoken language in audio recordings in order to verify their attitude toward Belgian Standard Dutch and tussentaal (other attitudinal studies can be found in Ghyselen 2009; Impe & Speelman 2007; Grondelaers et al. 2011; Grondelaers & Speelman 2013; Grondelaers & Van Hout 2016; Vancompernolle 2012). The research of Delarue (2016)

focused 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 stated that Colloquial Belgian Dutch has become increasingly common and accepted in various (mainly spoken) situations, whereas Belgian Standard Dutch loses its function in these contexts. These results suggest 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). As mentioned in Section 2.2., destandardization is defined by Coupland & Kristiansen (2011: 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 toward 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: 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). With regard to spoken language, Grondelaers & Van Hout (2011: 235) subscribe the gradual expansion of tussentaal to a process of 'endoglossic standardization', as the exoglossic standard (which is based on the Standard Dutch variety as used in the Netherlands) is only used in written registers and formal spoken language, but not in informal contexts and informal spoken language. Furthermore, the 'gradual abandoning of the VRT-norm and the absence of potential replacements' are suggestive of a process of destandardization, which could eventually lead to a 'standard language vacuum' (Grondelaers et al. 2011: 217). Jaspers & Van Hoof (2015), however, do not agree with this point of view, as they believe that the standard language ideal is still very much alive. They characterize the current Flemish language situation as a process of 'late standardization':

> The term late standardization alludes to the fact that standard and vernacular language use are currently being reconfigured, with the latter becoming more prominent in contexts and genres where the standard language used to be *de rigueur*, while the tension between standardizing and vernacularizing forces is intensifying and their relationship becoming more complex [...]. At the same time, conceptualizing the present situation as a case of late standardization captures the fact that this tension is not entirely new, and that in some contexts, vernacular

language use has always had (sometimes even considerable) presence (Jaspers & Van Hoof 2015: 36-37).

According to Ghyselen (2016), demotization is a more plausible scenario in the Flemish context, as the results of her study have shown that the standard language ideology is still alive among the informants. In certain formal contexts, the prestigious Standard Dutch variety is the preferred norm, whereas in other, more informal situations the dynamic tussentaal variety is considered to be the most appropriate norm. In addition, a perception study of Lybaert (2017: 112) indicates that language users evaluate Standard Dutch as 'the best and most correct variety, which can always be used but is especially appropriate for formal and official situations'. Tussentaal, on the other hand, is considered a 'neutral and even desirable variety' in informal situations, but also in certain formal situations, depending on the specific context (ibid.: 93). In brief, it is currently impossible to predict where these linguistic changes are heading, since Dutch linguists put forward different scenarios with regard to Flanders' linguistic future. As Ghyselen et al. (2016: 85) suggest, 'more attention has to be devoted to real or apparent time data which can actually demonstrate change'.

Unlike for spoken Dutch, the traditional standard language ideology has stayed intact right up until today for the written discourse. In the study of Delarue (2016), for instance, it was found that the language attitudes of Flemish teachers toward a written standard are much more explicit and pronounced compared to their attitudes toward spoken language. In other words, Flemish teachers regard the standard language norm primarily as a written norm. These results confirm that the language situation in Flanders is characterized by a linguistic dichotomy "between a formal (mostly written) Belgian Standard Dutch variety on the one hand, and an informal (mostly spoken) Colloquial Belgian Dutch one on the other" (Delarue 2016: 25). Most of the empirical research in the past has mainly focused, however, on spoken standard dynamics. Nevertheless, the influence of these spoken language changes on the written language has recently been tackled. Several studies have demonstrated, for instance, that Colloquial Belgian Dutch frequently occurs in online chat conversations of Flemish teenagers (De Decker 2014; Vandekerckhove 2007; Vandekerckhove & Nobels 2010). Given the recent linguistic dynamics in the language spoken on Flemish television, the question arises whether these processes also affect the subtitling practice in Flanders. To fill this gap, the present study will offer an insight into the influence of these (de)standardization tendencies on the subtitlers' linguistic choices.

2.3.3 Dutch on Flemish television

2.3.3.1 VRT and language planning

In paragraph 2.3.1, it was already mentioned that the media, and especially the Flemish public broadcasting corporation VRT, have played a significant language-planning role during the Dutch standardization process in Flanders (cf. Jaspers & Meeuwis 2006; Vandenbussche 2010; Van Hoof 2013; 2015 for more extensive analyses of VRT's language policy). Since its establishment in 1930, the Flemish public broadcaster claims to be one of the most important distributors of Belgian Standard Dutch in Flanders, carrying out its mission to serve as an example of correct language use toward its audience. From the 1970s onwards, VRT has employed a language advisor who ensures that the language that is used in radio and television programs does not deviate from the prevailing language conventions, viz. the use of Belgian Standard Dutch. Until 1996, VRT's language policy was explicitly oriented toward the northern norm, but with the arrival of the current language advisor Ruud Hendrickx, this conservative monocentric attitude regarding standard language has become more flexible, as Hendrickx (1998) recognized and tolerated explicitly the use of a *Belgian* variety of Standard Dutch. The language policy is written down in successive versions of VRT's *Taalcharter* [Language Charter] (Hendrickx 1998; 2007; 2012), in which is stated that the public broadcaster aims to be the norm for Belgian Standard Dutch, stimulating its presenters to speak standard language on the one hand, but leaving room for the expression of the Flemish community's identity and culture on the other hand. Right up until today, this mission seems to succeed, since Flemish speakers attach great normative value to the language that is used by the public broadcaster. As a result, Belgian Standard Dutch is often called VRT-Dutch by its users, referring to the language variety that is used in informative radio and television programs on the Flemish public broadcaster.

In the light of the increasing use of Colloquial Belgian Dutch in conversational and educational contexts (cf. 2.3.2; see also De Caluwe 2009; Delarue 2016), it does not come as a surprise that the recent linguistic developments in Flanders have also caused a shift in VRT's language policy. Initially, VRT's Language Charter (Hendrickx 1998) execrated the use of tussentaal, whereas more recent versions of this language policy (Hendrickx 2007; 2012) implicitly leave room for Colloquial Belgian Dutch: 'In fiction and comedy, VRT uses all varieties of Dutch' (Hendrickx 2012: 1). In these genres, the use of dialect and tussentaal is supposed to be functional, for instance, for a more authentic rendering of the interpersonal relations established between characters and the discursive situation (Hodson 2014; Lippi-Green 1997). In recent years, the actual language use on television has been studied, for instance, by Saman (2003), who has demonstrated that the use of Colloquial Belgian Dutch in radio spots increased between 1991 and 2001. In addition, Van Gijsel et al. (2008) have shown that language variation in radio and television advertising

depends on format, medium, and target public. In their study, the use of tussentaal significantly increased 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, in contrast to the requirements of the VRT Language Charter (Hendrickx 1998; 2012), TV hosts frequently speak tussentaal. Zenner et al. (2009) showed that participants in the reality show *Expeditie Robinson* adapted their language use to the conversational context, which implies that they were aware of the informal nature of Colloquial Belgian Dutch. In a diachronic study, Van Hoof (2013) demonstrated that together with its educational role, the use of Belgian Standard Dutch in television fiction has decreased in a period of thirty years, whereas the use of tussentaal has increased.

2.3.3.2 VRT's subtitling policy: theory and practice

The aforementioned studies have indicated that spoken language on Flemish television nowadays shows a lot of variation, which strongly varies according to the program genre or the characters (cf. Coupland 2010; 2014). Next to Belgian Standard Dutch, Colloquial Belgian Dutch is frequently used by TV hosts, and in commercials and entertainment programs. To ensure that the use of non-standard language varieties does not hamper the intelligibility, VRT intralingually subtitles its television programs. VRT attaches great importance to the compilation of its own subtitles, rather than buying subtitles from translation companies, like commercial stations tend to do. Initially, VRT provided open intralingual subtitles if the language variety spoken by the characters was considered unintelligible for the audience. Previous studies of Vandekerckhove et al. (2006; 2007; 2009) and Remael (2008) have demonstrated that this subtitling practice seemed to 'reflect the existence of shifting linguistic norms in Flanders' (Vandekerckhove et al. 2009: 609). On the one hand, Netherlandic fiction programs (e.g. *Baantjer*) were systematically provided with intralingual subtitles on Flemish television, which suggested that program makers believed that the Flemish audience had become alienated from the northern variety and was therefore in need of subtitles to understand the Netherlandic-Dutch speech. On the other hand, Flemish fiction series were often not subtitled, although a lot of Colloquial Belgian Dutch is used in these programs. This tendency was considered a reflection of the increasing tolerance toward the use of Colloquial Belgian Dutch on Flemish television. In other words, program makers assumed that tussentaal had become a commonly used, intelligible colloquial variety for the general Flemish audience. In non-fiction programs, however, the subtitling practice was more in line with the official language policy. The informative nature of this genre probably explains why colloquial speech was frequently subtitled in these programs. Nowadays, however, the Flemish public broadcaster no longer has complete authority on the use of open intralingual subtitles, as the program makers often decide whether a TV program is intralingually subtitled or not. Nonetheless, VRT provides closed intralingual

subtitles for all its programs, which the television viewers can retrieve through teletext (Hendrickx 2003b; 2011)[5].

Initially, there were no specific guidelines for intralingual subtitling and VRT sparingly used subtitles to avoid the stigmatizing effect they can create by representing the speaker as unintelligible (Hendrickx 2003b; 2011). Furthermore, intralingual subtitles were basically written in standard language (like interlingual subtitles) and this translation from dialect or tussentaal into Belgian Standard Dutch often affects the emotional value of the discourse (Hendrickx 2003b). Around the turn of the century, the public broadcaster developed an official subtitling policy, which was largely oriented toward the general well-being of its audience. In 2000, VRT set up a large-scale qualitative study to examine the needs and desires of its viewers. The results of this inquiry were published (Doens 2000) and used for the compilation of the first *Stijlboek voor Teletekstondertiteling* [Style Guide for Teletext Subtitling] (Dewulf & Saerens 2000), a book with guidelines for closed subtitling on the public broadcaster. First, it was found in this study that the hearing audience often evaluates open subtitling as disturbing, especially when the subtitles deviate too much from the spoken source tekst (Hendrickx 2011). In addition, Vandekerckhove et al. (2007) found that their informants disagreed about the necessity of intralingual subtitling, as the number of informants who declared that subtitles were disturbing was almost equal to the number of informants who indicated that subtitles were useful. When presented with unsubtitled fragments, there were almost as many informants who did not ask for subtitling as informants who actually required it. A transition from open to closed subtitles offered a solution for this division. Nowadays, viewers can choose whether they want to watch the program with or without subtitles, as Dutch-spoken programs are generally provided with closed intralingual subtitles, which can be retrieved through teletext. 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 2003b; 2011).

VRT's study also revealed that the deaf and hard-of-hearing find it absolutely necessary to have subtitles while watching television (Doens 2000). VRT is therefore their preferred broadcasting station, since almost every television program is subtitled. Moreover, they declared that non-standard utterances should be rendered as such in the subtitles, not only to retain the authenticity of the program, but also to make the subtitles match the mouth image of the characters. Furthermore, many deaf and hard-of-hearing viewers consider these subtitles a medium to learn the Dutch language in all its varieties

---

[5] At the time this dissertation is printed, the subtitling policy of VRT has changed. The subtitling department recently lost its authority over which TV programs are subtitled and which are not. Nowadays, these decisions are made by the channels and they tend to provide all fiction programs with open intralingual subtitles, with the exception of the soap and *Thuis* and comedy programs.

(Doens 2000; Slembrouck and Van Herreweghe 2004). However, this desire contravened VRT's norm-adherent language policy, in which the public broadcaster adopts an attitude of what Rosa (1994; 2001) names 'centralization', i.e. translating non-standard language into standard language, while ignoring the features of spoken verbal language. According to Rosa (2001), this strategy of centralization is mostly adopted by public broadcasters in an attempt to uphold the standard variety and to diffuse the prestige of the written norm (see also Cavalheiro 2008; Ramos Pinto 2009). Finally, VRT's subtitling department reached a compromise, of which the results were written down in the Style guide for Teletext Subtitling (Dewulf & Saerens 2000). Next to instructions regarding the position and layout of the teletext subtitles, this manual contained a limited number of guidelines concerning the use of standard and non-standard, colloquial language. According to this style guide, subtitlers were expected to translate the colloquial speech of the television program into standard language, thereby going against the desires of the hearing impaired audience. Nevertheless, VRT occasionally left room for colloquial lexical items, which could be '"more or less" reproduced in the subtitles.' Morphosyntactic tussentaal constructions, such as 'the colloquial forms of the personal pronoun (*ge/gij/u(w)* ['you(r)']) and the flexion of articles (e.g. *nen* ['a(n)']) and pronouns (e.g. *mijnen* ['my']) were not reproduced' [my translation] (Dewulf and Saerens 2000: 35)[6]. Furthermore, a non-standard lexical item was only used in the subtitles if all editors agreed that the colloquial lexeme was acceptable.

In 2009[7], a revised version of the subtitling style guide was published with more detailed guidelines concerning the use of colloquial language in closed subtitles. With regard to the lexicon, this new style guide prescribes that tussentaal is to be reproduced 'as much as possible' in the subtitles of fiction series and comedy to retain the 'couleur locale' of both the program and the characters (VRT 2009: 14), which suggests that VRT had become more tolerant toward the use of colloquial variants in the closed subtitles of certain television genres. Furthermore, if the TV host uses tussentaal, although he is expected to speak Belgian Standard Dutch, his colloquial words are to be reproduced in the subtitles. Colloquial morphosyntactic constructions, on the contrary, are to be replaced by their Belgian Standard Dutch alternative, although there are a few exceptions. The personal pronoun (example 1), incorrect articles (example 2), imperatives ending on *-t* (example 3), double use of *gaan* (example 4), proper names in combination with an article (example 5) and the use of *voor* or *van* instead of *om* (example 6) should always be translated into standard language[8]:

---

[6] With regard to open interlingual subtitles, VRT aims for the use of standard language.

[7] In 2017, VRT has published the newest version of its subtitling guidelines. This style guide was not included in this dissertation, since our corpus materials date from 2000 to 2016.

[8] The following examples are extracted from the corpora that were used for this research.

1. Wel, met hoe meer **ge** zijt, hoe sterker. (original speech)
   Met hoe meer **je** bent, hoe sterker. (subtitle)
   The more **you** are, the stronger. (translation)

2. Zo kan **de pond** wel eens fors in waarde zakken. (original speech)
   Zo kan **het pond** fors in waarde zakken. (subtitle)
   In this way, **the pound** can significantly decrease. (translation)

3. **Drinkt** uw glas leeg. (original speech)
   Drink je glas leeg. (subtitle)
   Drink up your glass. (translation)

4. Ik **ga** volgende week naar de markt **gaan.** (original speech)
   Ik **ga** volgende week naar de markt. (subtitle)
   I will **go** to the market next week. (translation)

5. Luc, 't is **de Frank**, uw broer. (original speech)
   Luc, het is **Frank**, je broer. (subtitle)
   Luc, it's **Frank**, your brother. (translation)

6. Ik heb dat nodig **voor** te kunnen slapen. (original speech)
   Ik heb dat nodig **om** te kunnen slapen. (subtitle)
   I need this to sleep. (translation)

The flexion of adjectives (example 7), pronouns (example 8) and articles (example 9) are not reproduced in the subtitles, except for the possessive pronoun *ons/onze* (example 10):

7. Was het bij u ne **mooien** baby? (original speech)
   Was het bij jou een **mooie** baby? (subtitle)
   Did you have a **pretty** baby? (translation)

8. Ik heb nog is gebeld op **zijne** gsm. (original speech)
   Ik heb nog eens gebeld op **zijn** gsm. (subtitle)
   I have called him on **his** mobile phone once again. (translation)

9. 't Is **ne** rare. (original speech)
   Het is **een** rare. (subtitle)
   It's **a** weird guy. (translation)

10. **Ons moe** zegt da altij. (original speech)
    Ons moe zegt dat altijd. (subtitle)
    Our mom always says that. (translation)

Also diminutives ending on *-ke* are converted into the Belgian Standard Dutch alternative ending on *-je*, except when referring to persons (example 11 and 12):

11. **Pake**, wees voorzichtig. (original speech)
    Pake, wees voorzichtig. (subtitle)
    Daddy, be careful. (translation)

12. Kom eens hier, **Tommeke**. (original speech)
    Kom eens hier, **Tommeke**. (subtitle)
    Come here, **Tom**. (translation)

Contrary to the style guide of 2000, the version of 2009 allows the semi-phonetic transcription of certain dialect words if it does not impede the intelligibility of the conversation. In other words, VRT has become more tolerant toward the use of Colloquial Belgian Dutch in subtitling nowadays. Not only does the style guide of 2009 allow more colloquial lexemes, there are also more concrete guidelines regarding the use of colloquial morphosyntactic constructions. With respect to open intralingual subtitling, guidelines are written down in the style guide *Normen en instructies voor open ondertiteling* [Norms and instructions for open subtitling], even though these guidelines are largely in line with the manual for teletext subtitling. Interlingual subtitles, however, are always written in Belgian Standard Dutch.

Although VRT's Style guide for Teletext Subtitling advocates the use of colloquial lexicon in closed intralingual subtitling to retain the authenticity of the television program and the characters, colloquial words are not consistently retained in their colloquial form in the subtitles. Remael et al. (2008), showed that the subtitlers' choices are rather arbitrary. Sometimes, unmistakable tussentaal words (e.g. *content* ['happy'], *zeveraar* ['driveller']) were reproduced in the subtitles, whereas other, commonly used lexemes (e.g. *gazet* ['newspaper'], *iedere* ['every'], *luidop* ['out loud']) were converted into Belgian Standard Dutch. Furthermore, they demonstrated that only a limited number of Colloquial Belgian Dutch words were used, although the subtitlers claimed that it was not their priority to translate the colloquialisms into standard language variants. According to a diachronic study of De Ridder (2015), the shift in VRT's language policy did not cause an evolution in the language that is used in the subtitles. Although the public broadcaster has become more tolerant toward the use of Colloquial Belgian Dutch in fiction programs, the number of Belgian-Dutch colloquialisms in the subtitles has not increased significantly in the past twenty years: interlingual subtitles are still written in Belgian Standard Dutch; intralingual subtitles in Flemish crime series contain practically as much

Belgian-Dutch colloquialism as two decades ago. Nonetheless, the aforementioned studies merely focused on Colloquial Belgian Dutch lexicon, omitting colloquial syntactic and morphological constructions, and they did not investigate the Flemish spoken source text and its potential influence on the subtitlers' linguistic choices. Furthermore, neither was the language used in the subtitles compared to the language used in other written translations and non-translations, nor was the effect of extralinguistic contexts (e.g. program genre) examined, since only one television genre was taken into consideration. To fill this gap, the present dissertation 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*; definitions of each genre will be given in Section 3.3.1). In addition, the subtitlers' linguistic choices will be compared to the linguistic choices of translators and authors of original Dutch texts. The outcome of this study will subsequently give us more insight into the current status and acceptance of Colloquial Belgian Dutch in Flanders.

Chapter 3
Research methodology

The present chapter will outline the methodology that was used for the investigation of norm-related linguistic variation in subtitling on Flemish television. This methodology is corpus-based, with a strong emphasis on exploratory multivariate statistics. In order to achieve the research goals and to find an answer to the research questions described in Chapter 1, three corpus-based case studies are carried out. This corpus-based approach is considered highly suitable for investigating subtitlers' linguistic choices in different subtitling contexts. A similar five step methodology is used in these three case studies: (i) formulation of the hypotheses, (ii) selection of the variables and data extraction, (iii) manual validation and annotation of the extracted corpus data, (iv) statistical analysis and visualization of the data in two-dimensional plots, which allow us to interpret the results and to draw conclusions with regard to the initial hypotheses. In case study 3, an additional research step will be taken, by performing a qualitative analysis of the results obtained in the quantitative study. In the following sections, these research steps will be discussed in more detail.

3.1   Corpus materials

To collect our data, we consulted different corpora. On the one hand, two already available 'general' corpora were used. Kennedy (1998: 19-20) defines a general corpus as a corpus 'assembled simply to make available a text base for unspecified linguistic research'. Such a corpus is 'typically designed to be balanced by containing texts from different genres and domains of use including spoken and written, private and public' (ibid). Both the SoNaR Corpus (3.1.1) and the Dutch Parallel Corpus (3.1.2) can be considered general corpora. It is important to bear in mind, however, that the SoNaR

Corpus is a 'dynamic' corpus to which new text material can be added. As a consequence, the use of updated versions of this software program could result in different frequency counts. Further refinement of the web tool OpenSoNaR may cause more or fewer hits for the same search query in future editions of the tool. For our case studies, we have used version 1.0. On the other hand, we built our own 'specialized' corpus (3.1.3), defined by Baker (2006: 147) as a corpus that is 'designed for a particular research project'. Like the Dutch Parallel Corpus, this corpus is a 'static' corpus, consisting of a 'static collection of texts [viz. subtitles, my comment] selected in some principled way, intended to be typical of the whole language or an aspect of the language at a particular time' (Kennedy 1998: 60). In other words, no future data will be added to the corpus, as it was exclusively built for this study.

Depending on the research objective(s) we want to achieve, we consulted one (or two) of these corpora. In case study 1, we extracted our data from both the SoNaR Corpus and the Dutch Parallel Corpus in order to compare the subtitlers' norm-related linguistic choices to the choices made in written translations and non-translations. The data of case study 2 were extracted from the SoNaR corpus only, as we zoomed in on the linguistic behavior of the subtitlers. By using this large corpus of VRT subtitles, we were able to investigate (i) whether the subtitles contain more colloquial lexemes than colloquial grammatical constructions and (ii) to what extent the subtitlers' linguistic choices are influenced by different contextual factors (e.g. *source language, program genre, speaker type*). In case study 3, we compared the subtitlers' linguistic choices to the original speech in the television program. Since the SoNaR Corpus does not contain the spoken television fragments, we had to build our own parallel corpus, consisting of the orthographic transcriptions of the spoken language in twenty Flemish television programs aligned with the corresponding intralingual closed subtitles. In the following sections, each corpus will be discussed in detail.

3.1.1   The SoNaR corpus

To explore the frequency distributions of Belgian Standard Dutch and Colloquial Belgian Dutch words and constructions in various subtitling contexts, we consulted the SoNaR Corpus[9], a 500-million word balanced reference corpus for contemporary (1954-present) written Dutch (Reynaert et al. 2010). The SoNaR Corpus was built on the initiative of the STEVIN programme, a Flemish/Dutch human language technology research programme, and funded by the Dutch and Flemish Government with the intention 'to serve as a general reference for studies involving language and language use' (Oostdijk et al. 2013:

---

[9] The acronym SoNaR stands for STEVIN Nederlandstalig Referentiecorpus, i.e. STEVIN Dutch Reference Corpus.

221). Not only is the corpus regionally stratified (Belgian Dutch vs. Netherlandic Dutch), it is also stylistically stratified across 36 text types, including newspapers, reports, emails, text messages, and subtitles. All texts (except for texts from social media such as Twitter, Chat, and SMS) have been tokenized, tagged for part of speech and lemmatized to allow for more efficient search queries (e.g. lemma searches to obtain all conjugation forms of a given verb). A disadvantage of the SoNaR Corpus, however, is that the part of speech tagging was done automatically, without a manual validation. As a consequence, certain words were incorrectly classified (e.g. the noun *gebruik* 'use' is often tagged as a verb), which can yield a biased result if one searches by word class.

In 2014, the web tool *OpenSoNaR*[10] was hosted online, which allows for analyzing and searching the large scale SoNaR Corpus by means of an online interface. One of the limitations of the online application is that the number of hits shown in OpenSoNaR is limited to 8,000,000. In other words, if the results of a query exceeds this limit, only the first 8,000,000 hits will be shown. This restriction did not cause any obstacles in our research, as the total number of attestations of the language variants did not exceed 10,000. Furthermore, the corpus does not contain the original source texts and also the meta information of the included texts is limited.

Figure 1 shows an example of a search query that was carried out in the OpenSoNaR interface. In the *Verken* ('Exploration') interface, the corpus distributions for each text type can be consulted, statistics from sub-corpora can be requested, n-grams from sub-corpora can be retrieved, and specific documents can be found by using the SoNaR document ID. The *Zoek* ('Search') interface offers four different search strategies: *simpel* ('simple'), *uitgebreid* ('extended'), *geavanceerd* ('advanced'), and *expert* ('expert'). A simple search query for the colloquial variant *gazet* 'newspaper', for instance, resulted in 4,141 hits in the entire SoNaR Corpus, of which 208 instances were found in the component with subtitles.

---

[10] https://portal.clarin.inl.nl/opensonar_whitelab/page/search
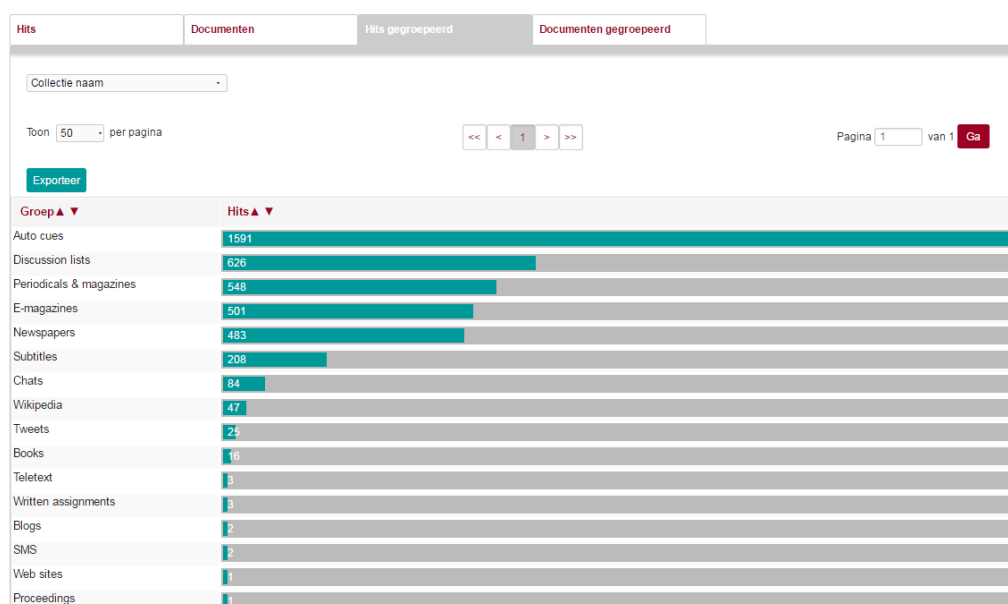
Figure 1. A query example that was carried out in the OpenSoNaR interface

Beside VRT's professional subtitles, one third of the subtitle material in the SoNaR Corpus comes from the OPUS Corpus compiled by Tiedemann, consisting of movie subtitles that were created by non-professionals (also called *fansubs*). For the purposes of this research, we obviously only selected the corpus component with subtitles created by the Flemish public broadcaster. This subcorpus consists of more than 18 million words (n = 18,687,891) and contains closed intralingual (i.e. source language is either Belgian Dutch or Netherlandic Dutch) and open interlingual (i.e. source language is English) subtitles of 109 TV programs that were broadcast by VRT[11] (on the channels *Eén* and *Canvas*) between 2000 and 2005. Table 1 gives an overview of the television programs that are included in the SoNaR Corpus.

---

[11] VRT confirmed that all subtitlers were Flemish by mother tongue.

1	1000 Zonnen en Garnalen	56	In vino
2	16+	57	Kaat en Co
3	1 jaar gratis	58	Karrewiet
4	Afrit 9	59	Kemphanen
5	Alfa Papa Tango	60	Kinderen van Dewindt
6	Alles voor de show	61	Kijk Uit!
7	America's Castles	62	Kokteel
8	Bal Mondial	63	Kongo
9	Ben en de Belgen	64	Kulderzipken
10	Blanke koning, rood rubber, zwarte dood	65	Kwesties
11	Blokken	66	Mijn Gedacht
12	Brussel Nieuwsstraat	67	Napels Zien
13	Buiten de Zone	68	National Geographic
14	Bureau voor Grote Vragen	69	Onder één dak
15	Confidenties in de Provence	70	Ooggetuige
16	Dag Sinterklaas	71	Op gelijke voet
17	De Avonturen van Kuifje	72	Op stap met dinosauriërs
18	De Boerenkrijg	73	Overleven
19	De Bovenste Plank	74	Pappenheimers
20	De Hopeloze Heks	75	Panorama
21	De Jaren van Verstand	76	Pas Geverfd
22	De Kat	77	Pluk de dag
23	De Keuze van Alloo	78	Rare Streken
24	De laatste dagen van Hitler	79	Recht op Recht
25	De Leukste Euw	80	Rubriek 700
26	De Mol	81	Schaduw van het Kruis
27	De Rode Loper	82	Schokkend Wereldnieuws
28	De Vloek van Vlimovost	83	Spike
29	De Wereld is Klein	84	Spring
30	De Wereld van Tarantino	85	Stafkes Straffe Kost
31	Dieren in nesten	86	Stille Waters
32	Dierenkliniek	87	Stories
33	Docu	88	The Millennial Pope
34	Don't drop the coffin	89	Thuis
35	En daarmee basta!	90	't Is proper
36	Eurosong for Kids	91	Quix
37	FC De Kampioenen	92	Spots
38	Flikken	93	Ten Huize Van
39	Geert Hoste shows	94	The 1940's house
40	Gentse Waterzooi	95	The Way to the Final
41	Getuige gezocht	96	Urbain
42	Grenzeloze Liefde	97	Verborgen Verleiders

43	Groot Licht	98	Vlaanderen Vakantieland
44	Halleluja!	99	Vriend of Vijand
45	Helden	100	Vrolijk Vlaanderen
46	Heterdaad	101	Vrouwengevangenis
47	Het ABC van de VRT	102	W817
48	Het Derde Oog	103	Weg wegens Wissel
49	Het Eiland	104	Wij, heren van Zichem
50	Het Leven Zoals Het Is	105	Windkracht 10
51	Histories	106	Witse
52	Huis Zoekt Date	107	Witte Raven
53	Ieder Zijn Wereld	108	Zalm voor Corleone
54	Ik zie je graag	109	Zonnekinderen in Schotland
55	In het spoor van de dino's		

Table 1. List of the television programs in the SoNaR Corpus

The data of both the first and the second case study were extracted from the SoNaR Corpus. However, at the time we collected the data for case study 1, the web tool OpenSoNaR was not yet publicly available. As a consequence, we had restricted access to the corpus, so our study was limited to the investigation of norm-adherent language use in 13 television programs[12], transmitted between 2001 and 2005 and provided with open interlingual or closed intralingual subtitles. The subtitles were delivered to us in XML format by a co-worker of the SoNaR research project and by means of a custom search engine, we extracted the data from this subcorpus. The total number of words in this corpus component is somewhat more than two million (n = 2,048,480).

3.1.2 The Dutch Parallel Corpus

To compare the linguistic choices in the subtitles of the SoNaR Corpus to the linguistic choices made in regular written translations and non-translations (cf. case study 1), we also consulted the Dutch Parallel Corpus (DPC; Macken et al. 2011). DPC is the result of the COMURE project (2006-2009) that was funded by the Dutch Language Union as part of the STEVIN programme. It is a bidirectional parallel translation corpus with (Belgian and Netherlandic) Dutch as a source language (translated into French or English) and as a target language (translations from French or English). It consists of more than 10 million words, it is stratified across 5 genres (literature, journalistic texts, administrative texts, instructive texts and texts for external communication), and it is sentence-aligned, part of speech-tagged and lemmatized. For this study, we only selected translations and non-translations that were published in Flanders, and eliminated the texts of which the source

---

[12] Viz. n° 7, 30, 33, 34, 48, 50, 51, 65, 75, 78, 82, 95, 101 in Table 1.

language was unknown. Table 2 gives an overview of the structure and size of the selected corpus components (total size: n = 3,778,854).

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

Table 2. Overview of the structure and word count of the selected component of the DPC

The entire corpus is released as full XML files and via a web interface[13] that was developed at Ghent University. This online tool supports simple and complex search queries, and presents the results as parallel concordances. Figure 2 shows an example of a search query that is carried out in the DPC interface. A search query for the variant *krant* 'newspaper', for instance, resulted in 87 hits in the Belgian-Dutch (NL-BE) subcorpus, of which merely the first 50 instances are shown in preview.

---

[13] http://dpcserv.ugent.be/comure

Figure 2. A query example that was carried out in the DPC interface

3.1.3 Specialized parallel corpus

In order to examine to what extent the original footage of the television program influences the linguistic choices of the subtitlers (cf. case study 3), we built a parallel corpus, containing the orthographic transcriptions of the spoken language in twenty Flemish television programs aligned with the corresponding intralingual closed subtitles[14]. These programs were aired by the public broadcaster VRT between 2014 and 2016[15], and they were selected in function of the genre classification that was made in case study 2[16]. As a result, each genre contains four different television programs of which

---

[14] The corpus is available on request or it can be consulted online: http://www.eqtis.ugent.be/lynn_prieels/

[15] We are aware that the TV programs in this corpus were broadcast more than ten years later compared to the TV programs in the SoNaR Corpus, which could have an effect on the subtitlers' linguistic choices. Nevertheless, the data were kept separated in the analyses to ensure that the results were reliable for a certain period. In the conclusions of Chapter 7, we will elaborate in this.

[16] The analyses in case study 2 are based on 5 genre categories: *children's television*, *comedy*, *documentaries*, *fiction*, and *light entertainment*. In section 3.3.1, this genre classification will be discussed in detail.

in each case 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 this corpus. Section 3.3.1 discusses how this genre classification was made.

Genre	TV program	Episode
Documentaries	*De vrije markt*	18/06/2016
		25/06/2016
	De Zevende Dag	05/06/2016
		12/06/2016
	Greece, the islands	Episode 1 (20/06/2016)
		Episode 2 (21/06/2016)
	Koppen	Operatie Vigilant Guard (01/06/2016)
		De zwarte lijst (08/06/2016)
Fiction	*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)
	Thuis	Season 21, episode 4008 (01/06/2016)
		Season 21, episode 4013 (08/06/2016)
	Tom & Harry	Season 1, episode 9 (06/04/2015)
		Season 1, episode 10 (13/04/2015)
Children's TV	*D5R*	Season 3, episode 31 (01/06/2016)
		Season 3, episode 32 (02/06/2016)
	Helden	Season 3, episode 2 (04/06/2016)
		Season 3, episode 3 (11/06/2016)
	Klein gespuis	Season 1, episode 5 (30/05/2016)
		Season 1, episode 7 (1/06/2016)
	Karrewiet	Episode 107 (30/05/2016)
		Episode 109 (01/06/2016)
Comedy	*Achter de feiten*	Episode 9 (14/05/2014)
		Episode 10 (21/05/2014)
	Echt niet ok!	Episode 8 (09/02/2016)
		Episode 9 (16/02/2016)
	Hoe is het zover kunnen komen?	(19/06/2016)
		(29/06/2016)
	Nieuw Texas	Episode 6 (03/06/2015)
		Episode 7 (10/06/2015)
Light entertainment	*1000 Zonnen*	(31/05/2016)
		(01/06/2016)
	Blokken	Episode 4694 (30/05/2016)
		Episode 4695 (31/05/2016)

	Dagelijkse kost	Season 6, episode 1536 (30/05/2016)
		Season 6, episode 1537 (31/05/2016)
	Over eten	(01/06/2016)
		(08/06/2016)

Table 3. List of the television programs in the specialized parallel corpus

Each transcription was given a unique code, depending on the transcribed mode, the program genre, the particular TV program, and the episode. Table 4 contains the specification of each code that was used for the codification of the transcriptions. For instance, transcription SUD2.1 is a transcription of the subtitles (SU) of the first episode (1) of the documentary (D) *De Zevende Dag* (2).

	Code	Explanation
Mode	SP	Spoken
	SU	Subtitles
Genre	CT	Children's television
	C	Comedy
	D	Documentaries
	F	Fiction
	LE	Light entertainment
TV program	1	First program in the genre category
	2	Second program in the genre category
	3	Third program in the genre category
	4	Fourth program in the genre category
Episode	1	First episode of the TV program
	2	Second episode of the TV program

Table 4. Overview of the codes that were used to encrypt the transcriptions in the specialized parallel corpus

3.2   Variable selection

In order to trace linguistic norm adherence in various translated and non-translated genres, we chart the frequency distributions of Belgian Standard Dutch (BSD) words and constructions in contrast to their Colloquial Belgian Dutch (CBD) counterparts in our corpora. The basic idea underlying this approach is that norm adherence can only be reliably measured if the proportion of BSD features is studied in combination with the proportion of their CBD alternatives. This is what Speelman et al. (2003) called the profile-

based approach, with each combination of a BSD variant (e.g. *een beroep doen op* 'to make an appeal to') and a CBD variant (e.g. *beroep doen op* 'to make an appeal to') that cover the same meaning or linguistic function being a profile. In this context, a profile is based on the concept of the sociolinguistic variable in the field of classic variational linguistics (e.g. Labov 1966). This usage-based technique does not investigate which words and constructions belong to a certain language system or language variety, but it verifies which linguistic features are actually used and how often these are used. Furthermore, the language variants are seen as subtypes within the same profile, thus referring to the same concept. As a result, these words and constructions are always investigated in relation to synonymous words and constructions, rather than being regarded as autonomous, isolated entities. One of the main characteristics of the profile-based method is the exclusion of referential ambiguity. For instance, the Dutch variant *bot* has different meanings, viz. 'boot' vs. 'bone'. As the profile-based approach implies that each variant in the profile should be able to replace its linguistic counterpart(s), attestations that not fulfil this condition will be removed from the data set. More advantages of this technique in the context of Translation Studies can be found in De Sutter et al. (2012a) and Delaere & De Sutter (2013). For each case study, we selected lists of language features in function of the research goals.

3.2.1   Case study 1

To compare the subtitlers' norm-related linguistic choices to the choices made in written translations and non-translations, we selected 8 linguistic profiles, which are presented in Table 5. 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 research 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: 1), we have chosen to adopt this terminology. Furthermore, General Standard Dutch features are automatically considered standard language in Belgium.

The list of profiles in Table 5 is almost identical to the list in Delaere et al. (2012), where the norm-adhering hypothesis was studied for written translations vs. non-translations in isolation. The selection of the profiles in Delaere et al. (2012) was based on a number of normative sources, such as *Taaladvies*, the database on language usage provided by the Dutch Language Union, which is currently considered the most important source of normative guidelines for language users in the entire Dutch language area, and *Correct Taalgebruik* [Correct language use] (Penninckx et al. 2001). The profiles were only selected if these sources agreed in characterizing the variants as BSD or CBD. Furthermore, to ensure that the profile set would be able to 'visualize the dispersion of standard language

versus non-standard language without being distorted by other factors' (Delaere 2015: 120), the profiles were only selected if the variants differed according to their normative status (standard language vs. non-standard language), and not because of their formality level or regional character. Because of this strict approach, the selection process resulted in a relative small set of relevant profiles.

As the first case study is partially based on the data of Delaere (2012), we also adopted this research design, including the small profile list. Moreover, as the normative status of two Belgian-Dutch lexemes has currently changed, we have chosen to eliminate two profiles from the original list in Delaere et al. (2012) (*geraken-raken* 'to get'; *bekomen-verkrijgen* 'to obtain')[17]. In addition, the orthographic profiles *een van de-één van de* ('one of the'), *te veel-teveel* ('too much'), and *tenminste-ten minste* ('at least') were also deleted, since *Taaladvies* labels the deviant forms as incorrect, rather than as CBD.

Profile	Belgian Standard Dutch	Colloquial Belgian Dutch	Translation or meaning
1	*akkoord gaan met*	*akkoord zijn met*	*to agree with*
2	part + aux + inf aux + inf + part	aux + part + inf	position of the participle in the verbal end group
3	*zulke* + meervoud	*zo'n* + meervoud	*such* + plural noun
4	*een beroep doen op*	*beroep doen op*	*to make an appeal to*
5	*proberen te* + inf	*proberen* + inf	*to try (to)* + inf
6	*op het eerste gezicht*	*op het eerste zicht*	*at first sight*
7	*beginnen te* + inf	*beginnen* + inf	*to start (to)* + inf
8	*zodra*	*van zodra*	*as soon as*

Table 5. List of the linguistic profiles used in case study 1

3.2.2 Case study 2

The second case study zooms in on the linguistic choices made by Flemish subtitlers. To verify whether subtitles on Flemish television contain more colloquial lexemes than colloquial grammatical constructions, we extended our initial profile set of the first case study and divided these linguistic profiles into three sets: a set with lexical-paradigmatic profiles (cf. Table 6), a set with constructional-paradigmatic profiles (cf. Table 7), and a

---

[17] At the time this dissertation is printed, the normative status of *akkoord zijn met* has also changed into Belgian Standard Dutch.

set with syntagmatic profiles (cf. Table 8). For the compilation of these lists, we consulted several normative sources. The lexical-paradigmatic profiles were first extracted from the *Referentiebestand Belgisch-Nederlands* [RBBN; Reference file Belgian Dutch] (Martin 2005), a collection of 4,000 typical Belgian-Dutch words and expressions. For this study, we only selected the colloquial variants that were labelled 'substandard'[18]. The RBBN was compiled in 1998 and updated in 2005 by the *TST-Centrale voor data en software van het Nederlands*. Next, Van Dale dictionary (Geerts and den Boon 1999; den Boon and Geeraerts 2005) was consulted to verify the status of each variant in the lexical profiles. The profiles were only selected if these normative sources agreed in characterizing the variants in each profile as standard language or colloquial ('*BE, spreektaal*' or '*BE, niet algemeen*' in Van Dale)[19]. Figure 3 shows, for instance, how the lexical-paradigmatic profile *fiets/rijwiel-velo* ('bicycle') was selected.

1.	RBBN	$velo_{CBD}$
		$fiets_{BSD1}$
2.	Van Dale (1999; 2005)	$velo_{CBD} = fiets_{BSD1}$
		$fiets_{BSD1} = rijwiel_{BSD2}$
		$rijwiel_{BSD2} = fiets_{BSD1}$
3	Lower limit	50 attestations in total for each profile

Figure 3. Selection process of a lexical-paradigmatic profile in case study 2

First, we selected the CBD variant *velo* and the BSD variant *fiets* from the RBBN. Next, both variants were looked up in Van Dale dictionary to find potential synonyms (here: *rijwiel*). Finally, we verified the presence of all variants in the SoNaR Corpus. In addition, we set a threshold with regard to two criteria. First, the variants were only retained if they occurred at least once in the corpus. The variant *weetgraag*, for example, was initially part of the profile *nieuwsgierig/benieuwd-curieus* ('curious'). However, this variant was eliminated, since it did not occur in the SoNaR Corpus. Second, the total number of attestations of all variants in the same profile had to be at least 50 in order to have a significant profile. Although *sterke drank-korte drank* ('liquors') was initially a profile candidate, the total sum of occurrences of both variants within this profile was merely

---

[18] Other labels that are used in the RBBN are, for instance, *vrije alternanten* 'free alternatives' and niet-gelexicaliseerde varianten 'non-lexicalized variants' (cf. manual RBBN on the website of the TST Centrale http://tst-centrale.org)

[19] VRT confirmed that the subtitlers also work in accordance with these labels.

41. As a consequence, this profile was not retained in the final set. With regard to the example above, this selection process resulted in a lexical-paradigmatic profile, consisting of three variants:

Status	Variant	Number of attestations
Colloquial Belgian Dutch	velo	61
Belgian Standard Dutch	fiets	1257
Belgian Standard Dutch	rijwiel	7

Furthermore, some lexemes turned out to be difficult to define, as Van Dale offered an unlimited number of synonyms. For example, the colloquial lexeme *zever* 'twaddle' has numerous BSD alternatives (e.g. *onzin, flauwekul, kletspraat, larie, apekool,...*). As a consequence, we decided to eliminate this lexeme, since we were not able to compile a well-defined profile. This selection process resulted in a set of 15 lexical-paradigmatic profiles of which an overview is given in Table 6. The lexemes in each profile belong to the same lexical-semantic paradigm, i.e. they have the same denotation.

Profile	Belgian Standard Dutch	Colloquial Belgian Dutch	Translation or meaning
1	*autosnelweg* *autoweg* *snelweg*	*autostrade*	*motorway*
2	*bestelwagen* *bestelauto*	*camionette*	*delivery van*
3	*fiets* *rijwiel*	*velo*	*bicycle*
4	*handtas*	*sacoche*	*handbag*
5	*jas*	*frak*	*coat*
6	*krant* *dagblad*	*gazet*	*newspaper*
7	*laars*	*bot*	*boot*
8	*motor* *motorfiets*	*moto*	*motorbike*
9	*nieuwsgierig* *benieuwd*	*curieus*	*curious*
10	*oom* *ome*	*nonkel*	*uncle*
11	*het platteland*	*de buiten*	*countryside*
12	*schrikken*	*verschieten*	*to be frightened*
13	*stropdas* *das*	*plastron*	*tie*
14	*vrachtwagen*	*camion*	*truck*

| 15 | *wastafel* | *lavabo* | *sink* |
| | *wasbak* | | |

Table 6. List of the lexical-paradigmatic profiles used in case study 2

The same selection process was applied for the grammatical profiles, although we used grammatical reference works, viz. *Taaladvies* and the *VRT-Stijlboek* (Hendrickx 2003a) instead of lexicographical works (viz. RBBN and Van Dale). We opted for these normative sources, since VRT's language policy and subtitling guidelines are also based on these reference works. With regard to the grammatical profiles, we made a distinction between constructional-paradigmatic and syntagmatic profiles. Table 7 shows the constructional-paradigmatic profiles, consisting of interchangeable constructions with the same meaning or function. Table 8 presents the syntagmatic profiles, which contain word order alternatives.

Profile	Belgian Standard Dutch	Colloquial Belgian Dutch	Translation or meaning
1	adj + *om* + *te* + inf	adj + *om* + inf	adj + *(to)* + inf
2	*beginnen te* + inf	*beginnen* + inf	*to start (to)* + inf
3	*een beroep doen op*	*beroep doen op*	*to make an appeal to*
4	*durven te* + inf	*durven* + inf	*to dare (to)* + inf
5	*mocht(en)*	*moest(en)*	hypothetical clause
	als		
6	*niet hoeven*	*niet moeten*	*not have to*
7	*op het eerste gezicht*	*op het eerste zicht*	*at first sight*
8	*over*	*na*	time indication
9	passief	passief +	passive clause
		geworden/geweest	
10	vz + vz + *en*	vz + vz	preposition +
			preposition
11	*proberen te* + inf	*proberen* + inf	*to try (to)* + inf
12	*zeker weten dat*	*zeker zijn dat*	*to be sure of*
	er zeker van zijn dat		
13	*(zo)als* + su	*(zo)als* + ob	*like* + object
14	*zodra*	*van zodra*	*as soon as*
15	*zulke* + meervoud	*zo'n* + meervoud	*such* + plural noun

Table 7. List of the constructional-paradigmatic profiles used in case study 2

Profile	Belgian Standard Dutch	Colloquial Belgian Dutch	Translation or meaning
1	part + aux + inf aux + inf + part	aux + part + inf	position of the participle in the verbal end group
2	NP + aux +inf	aux + NP + inf	position of the noun phrase in the verbal end group
3	PA + aux + inf	aux + PA + inf	position of the pronominal adverb in the verbal end group
4	part + inf + inf	inf + part + inf	position of the infinitive in the verbal end group

Table 8. List of the syntagmatic profiles used in case study 2

3.2.3 Case study 3

The main objective of case study 3 is to investigate to what extent the subtitlers' linguistic choices are influenced by the original footage of the television program. In this study, we will not adopt the profile-based approach, since we do not want to determine the frequency distributions of BSD words and constructions in contrast to their CBD counterparts. Instead, we want to calculate how often Flemish subtitlers reproduce the spoken colloquialisms from the original source text rather than converting these colloquial variants into BSD. Therefore, we selected a number of CBD items. Contrary to case studies 1 and 2, where we applied a top-down procedure to build the profile sets, the feature list in case study 3 is compiled by means of a bottom-up selection, starting from the subtitle corpus that was built for this case study (cf. 3.1.3). We have deliberately chosen to start from the subtitle corpus, since we are interested in which colloquial features are used in the subtitles in the first place. In the second place, we will verify whether these colloquialisms also occur in the spoken source text.

In a first step, the subtitles were manually scanned for colloquial features. To verify whether a variant was labelled CBD, we consulted Van Dale (2015), *Taaladvies* 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.

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 found in the transcriptions of SULE3.1, SULE2.1, and SUC2. 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 transcription of SPLE2.1. In the corresponding subtitles, 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.*

In total, 55 different CBD features were extracted from the corpus. These items were divided into three types: a set with lexical features (cf. Table 9), a set with morphological features (cf. Table 10), and a set with syntactic features (cf. Table 11).

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	*afbollen*	*get out*
7	*brol*	*trash*
8	*buizen*	*to flunk*
9	*camionette*	*delivery van*
10	*chance*	*luck*
11	*chapelure*	*breadcrumbs*
12	*chichi madam*	*chichi lady*
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*	*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*	*later*
32	*sjotten*	*play soccer*
33	*smossen*	*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*

Table 9. List of the lexical features that were used in case study 3

Feature	Colloquial Belgian Dutch	Translation
1	adjectief (+*e*)_fout	flexion of the adjective
2	bezittelijk vnw (+*e*)_fout	flexion of the possessive pronoun
3	diminutief *-ke*	diminutive
4	*ikke*	personal pronoun *I*
5	object *u*	object *you*

Table 10. List of the morphological features that were used in case study 3

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

Table 11. List of the syntactic features that were used in case study 3

One of the consequences of applying this methodology is that we only selected the colloquial features that were actually used in our subtitle corpus so 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 colloquial elements into standard language. 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 Gijsel 2008) in the past few decades. We selected 9 features that are considered typical ingredients of CBD and added them to our data set in order to investigate how often subtitlers convert these elements into BSD. An overview of these colloquial features is given in Table 12.

Feature	Colloquial Belgian Dutch	Translation
	Morphological features	
1	verbuiging lidwoord: *den, ne(n)*	flexion article
2	verbuiging aanwijzend vnw	flexion demonstrative pronoun
3	*ge/gij*	personal pronoun *you*
4	1$^{\text{ste}}$ persoon enkelvoud + *n*	deviant conjugation 1st person singular
5	2$^{\text{e}}$ persoon enkelvoud	deviant conjugation 2nd person singular
	Syntactic features	
6	*van/voor*+ beknopte bijzin	reduced clause
7	dubbele negatie	double negative
8	redundant *dat*	redundant *that*
9	subjectsreduplicatie	reduplication of the subject

Table 12. List of additional *tussentaal* features that were used in case study 3

3.3   Data extraction, manual validation and annotation

Using large corpora like the SoNaR Corpus and the DPC is highly suitable for the investigation of (norm-related) linguistic variation, because a large reference corpus increases the chance of retrieving relevant data. A disadvantage of such large corpora, however, is that a lot of language variants, and particularly lexemes, are polysemous or homonymous. As the profile-based approach implies that each variant in the profile should be able to replace its linguistic counterpart(s), all extracted data had to be manually analyzed in the specific context in which they occurred to verify that they were synonymous with and could be replaced by their BSD or CBD alternative respectively. If not, the attestation was considered irrelevant and removed from the data set. For example, the lexemes in the profile *laars* vs. *bot* can be used to refer to footwear, viz. a boot. Both variants have the same meaning and are thus interchangeable. However, the colloquial variant *bot* can also be used in the BSD meaning of 'bone'. If that was the case, this attestation was deleted from the data set. This validation process guaranteed that the statistical analysis would measure differences between the formal alternatives of a profile and not between their meanings. The data extraction and validation resulted in a final data set of 1,933 relevant instances for case study 1 (SoNaR: n = 887; DPC: n = 1,046); 36,470 validated observations for case study 2 (BSD: n = 27,068; CBD: n = 9,402); and 1,756 relevant CBD attestations (spoken: n = 1,616; subtitles: n = 140) and 1,490 relevant BSD attestations (in the subtitles) for case study 3.

In a next step, the validated data from the SoNaR corpus (cf. case study 1 and 2) were manually annotated for the contextual parameters *(program) genre*, *source language* and *speaker type*; the data from the DPC were already tagged for *genre* and *source language*, and the data from the new specialized corpus (cf. case study 3) were not annotated, because these data were only used to investigate which colloquial features were reproduced or translated in the subtitles, but not to measure the effect of contextual factors. For the SoNaR Corpus, the annotation of *source language* and *speaker type* was performed on the basis of the original footage. The former parameter turned out to be unproblematic: the source language was either Belgian Dutch, Netherlandic Dutch or English. With respect to the speaker type, the subtitles were subdivided in *voice-over* (i.e. either the presenter who introduces and concludes the program or the narrative voice who gives off-screen comments) or *actor/interviewee* (i.e. a dialogic context). Finally, the annotation of *program genre* had two phases: in the first case study, the television programs were divided in two general genres, whereas in case study 2, a more fine-grained genre classification was applied. The next section describes how this genre division was accomplished.

3.3.1 The genres

In case study 1, the annotation of the contextual parameter *genre* was partially based on the genre list in Van Gijsel et al. (2008: 210–212), which makes a distinction between journalistic television programs and entertainment programs. The main objective of journalistic programs is to inform the audience (e.g. a commentary, introduced by a presenter and/or commented on by a narrative voice), whereas the main objective of entertainment programs is to divert the audience (e.g. a docusoap or reality TV). This genre division was supposed to be suitable, as the proportion of the selected subtitle component of the SoNaR Corpus was rather limited (viz. *n* = 2,048,480) in the first case study. Table 13 shows the proportion of both genres in the corpus.

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

Table 13. Word count of the program genres in the corpus of case study 1

As can be seen, the distribution between journalistic and entertainment subtitles in the corpus is highly skewed: journalistic subtitles constitute almost 97% of the entire subtitles corpus component. Although this obviously is not an ideal distribution, it does not need to worry us too much. The results of the statistical analyses to be performed in the remainder of case study 1 are not affected by this skewed distribution, since the distances and associations between the different variables are based on their relative frequencies (and not their absolute frequencies).

As mentioned in Section 3.1.1, case study 1 is based on a subcomponent of the subtitle corpus in SoNaR (*n* = 2,048,480), whereas the entire subtitle component of the SoNaR Corpus was consulted in case study 2 (*n* = 18,687,891). As a consequence, the genre classification *journalism* vs. *entertainment* of case study 1 turned out to be too general when it was applied to the data in case study 2, so we divided the television programs into more specific genres. This new classification was largely based on the genre division of Creeber (2008), who anecdotally outlined the repertory of genres in the television landscape. He distinguished ten main genres, subdividing each of them into various subgenres (cf. appendix 1 for an overview of Creeber's classification, which will be discussed below). This classification, however, is not well-defined and the genre definitions are therefore not always mutually exclusive. Consequently, this approach does not fit our purposes, since we want to categorize each television program into a unique genre and use this genre classification as a key factor in multivariate analysis. As a result, we had to adapt Creeber's classification to our specific research needs. To ensure that our research purposes would not influence the classification, we defined three contextual subparameters: *main purpose, target audience* and *cast.* These parameters allow us to

identify and categorize different program genres independently of the response variable of this study, viz. the linguistic choices made by subtitlers. These parameters are comparable to what Biber & Conrad (2009) called the 'situational context', in which a certain register or language variety is used: 'linguistic features tend to occur in a register because they are particularly well suited to the purposes and situational context of the register' (Biber & Conrad 2009: 6). Table 14 gives an overview of the five program genres and their distinctive parameters (cf. appendix 2 for an explanation of the parameters and values in the table):

Program genre	Main purpose	Target audience	Cast
Fiction	entertaining	all ages	actors
Comedy	making laugh	all ages	actors
Children's television	infotaining	children	actors + non-actors
Light entertainment	infotaining	all ages	actors + non-actors
Documentaries	informing	all ages	actors + non-actors

Table 14. Overview of the genre classification used in case study 2

The left column in Table 14 contains the genres that were used to categorize the 109 television programs in our corpus. Compared with the genre division of Creeber (2008), our classification has been reduced to five program genres, since various genres (e.g. *costume drama*, *television news*, *music on television*) did not appear in the SoNaR Corpus. In a last step, we applied the interannotator agreement procedure (Nowak & Rüger 2010) in order to validate our classification. Two independent annotators were asked to categorize the television programs in the corpus into the five genres by using our classification table. This procedure had two phases: the first annotation round showed that some genres were not successfully defined, which resulted in a problematic classification of various TV programs. After having modified the criteria underlying the assignment of values, most of the television programs could be categorized successfully (K = 0.93)[20]. Only a few programs (e.g. the weather report, the King's Christmas speech) did not fit our classification, so they were left out. A brief description of the genres that were used in this study can be found below.

3.3.1.1 Fiction (drama)

Fiction programs are television programs of which the main purpose is to *entertain* the audience (the typical examples are *drama series* and *soaps*). These programs are intended

---

[20] K is the Cohen's Kappa coëfficient (Cohen 1968) which measures the inter-rater agreement. A K-score between 0.81 and 1.00 is interpreted as 'very good' or 'very strong agreement'.

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

3.3.1.2 Comedy

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

3.3.1.3 Children's television

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

3.3.1.4 Light entertainment

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

3.3.1.5   Documentaries

The main purpose of documentaries is to *inform* the audience (a typical example is a commentary, introduced by a presenter and/or commented by a narrative voice). These programs are intended for an audience of *all ages* and the cast consists of both *actors* (playing a role by using a script) and *non-actors* (spontaneous speech). Within this genre, Creeber (2008) distinguishes between *observational documentary* and *educational programming.* As Creeber mentioned, 'the provision of educational television has always targeted **children**' (Creeber 2008: 131; my emphasis), therefore we categorized the television programs of *educational programming* as *children's television.* Consequently, the remaining subgenre *observational documentary* was named *documentaries.*

3.4   Quantitative analysis: profile-based correspondence analysis

The results of this dissertation are mainly based on quantitative statistical methods that have recently become popular in Corpus-based Translation Studies. Furthermore, the application of multivariate techniques in this research domain has already yielded several valuable results (e.g. Delaere 2015; De Sutter et al. 2012a; Diwersy et al. 2014; Gries 2010; Rybicki 2012), and is therefore strongly stimulated (e.g. De Sutter et al. 2012b; Oakes & Ji 2012). The present study wants to contribute to this emerging trend by adding a multivariate dimension (viz. studying many linguistic features and many contextual parameters simultaneously) to the study of linguistic variability in AVT, as well as by linking the research results more explicitly to well-known explanatory principles in Translation Studies, viz. standardization or normalization (e.g. Delaere et al. 2012 for a discussion of the relationship between both principles). By investigating various contextual factors at once, this study will verify which factors have an influence on the linguistic choices of Flemish subtitlers, and how these factors are related to each other. Thanks to this multifactorial approach, we will also be able to generate results that are more widely applicable, and which are not based on the individual behavior of a single variable.

To verify how the linguistic profiles relate to the contextual parameters *source language*, *program genre*, and *speaker type*, we will apply an exploratory multivariate

statistical technique called profile-based correspondence analysis[21] (Plevoets 2008; 2015). This technique has proved to be a particularly suitable method for this type of research (see for example Delaere et al. 2012; De Sutter et al. 2012; Delaere and De Sutter 2013). Profile-based correspondence analysis is an extended version of standard correspondence analysis (Greenacre 2007) and differs from the latter in that it is made sensitive of the profile structure in our data set[22]. One of the strengths of this statistical technique is the spatial representation of the data in a two-dimensional plot. This statistical technique analyzes the associations between the rows (profiles) and the columns (contexts) of a frequency table to measure the relationship between these external parameters on the one hand, and the use of BSD and CBD on the other. Table 15 contains a section of the frequency table that was used for the analysis in case study 2, displaying the total number of attestations for every language variant in each program genre.

Variants	Label	Fiction	Children's television	Light entertainment	Comedy	Documentaries
jas	BSD	318	42	149	74	60
frak	CBD	40	0	1	8	2
nieuwsgierig	BSD	110	57	91	16	46
benieuwd		235	63	397	60	12
curieus	CBD	97	1	11	29	4
part + aux + inf	BSD	204	72	246	35	343
aux + inf + part		46	6	56	17	224
aux + part + inf	CBD	236	34	209	50	160
zodra	BSD	79	20	108	18	257
van zodra	CBD	28	2	44	6	15
...	...	...	...	...	...	

Table 15. Overview of the profile frequencies per program genre in case study 2

In a first step, the correspondence analysis calculates two matrices with distances: one for the distances between the rows (e.g. the association between the variants *jas* and *frak* 'coat' for the different program genres and source language varieties) and one for the distances between the columns (e.g. the association between the genre *fiction* and the genre *children's television* for all language variants). This calculation is based on the chi-square test. Second, the resulting distances are visualized in a two-dimensional plot by reducing the original, multidimensional matrices to two-dimensional matrices. The distances in these two reduced matrices are then rendered in a biplot (i.e. a type of

---

[21] All analyses in this dissertation were done in R (R Development Core Team, 2014), using the following command: dataset.crg <- corregp(variant ~ (genre + source.language + program purpose + target audience + cast)^2, data= data.read, b = 3000).

[22] The advantages of the profile-based approach were already described in section 3.2.

exploratory graph that generalizes the simple two-variable scatterplot), in which the distance between two data points indicates the exact nature of the association between them: the smaller the distance between the linguistic variants, the more closely they are related to each other (and vice versa). In other words, the relative distances between the data points and the way they are clustered determine the interpretation of the results. The fact that 'both the column and the row points [are projected] into the same subspace (biplot), thus allowing us to inspect the position of both observations and variables at the same time' is one of the greatest advantages of this technique (my addition; Jenset & McGillivray 2012: 317).

To verify whether the distances between the data points in the plot are statistically significant, confidence ellipses are calculated for each of the language varieties. These ellipses are the two-dimensional representations of the well-known confidence intervals (Reiczigel 1996) and were calculated through bootstrapping (3000 resamples). For every analysis, the confidence level was set at 95%, which means that it can be asserted with 95% certainty that the distances between the varieties is statistically significant ($p < 0.5$) if the confidence ellipses of two language varieties do not overlap. In this respect, it is important to note that the size of the ellipses is also negatively correlated to the number of corpus instances underlying these ellipses (thus representing relatively larger statistical uncertainty). The visualization of the associations between the variables and the contextual factors in a two-dimensional plot makes profile-based correspondence analysis highly suitable for an exploratory analysis to find patterns in our data.

3.5   Qualitative analysis

In case study 3, a qualitative dimension will be added to the quantitatively oriented methodology of this dissertation. 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. This qualitative analysis is based on semi-structured interviews, evaluation reports, and observational data that were collected at VRT's subtitling department. 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 goal of this qualitative study is to get more insight into the subtitlers' attitude toward 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 look into the guidelines by which VRT's subtitlers are particularly bounded. Next, we organized an interview with two subtitlers: the first interviewee (subtitler 1) is currently working at the editorial board of the T888-department (intralingual closed subtitling); the second interviewee (subtitler 2) is currently working at the editorial board of translations and interlingual subtitling. 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 toward the use of CBD in some corpus subtitles of this study. Each interview lasted about sixty minutes and was audiorecorded. 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 explained her choice for a specific colloquial of standard 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)

Chapter 4
Case study 1: analyzing linguistic norm adherence in Belgian-Dutch subtitles, regular written translations, and non-translations

The first case study compares the linguistic choices made in written translations and non-translations to the linguistic choices made in interlingual and intralingual subtitles in order to investigate to what extent subtitlers, translators and writers of original Dutch texts conform to the Standard-Dutch norm. This goal will be achieved by examining the distribution of Colloquial Belgian Dutch (CBD) variants and Belgian Standard Dutch (BSD) variants in both the SoNaR Corpus (for the subtitles) and the DPC (for translated and non-translated texts). In addition, we will also verify whether the contextual factors *source language*, *program genre*, and *speaker type* have an influence on the linguistic choices of the subtitlers.

4.1   Hypotheses

Based on the information in Chapter 2, we can formulate two hypotheses.

> ➢ **Hypothesis 1:** Flemish subtitlers are less norm-adhering than translators of regular written texts, i.e. subtitlers use fewer BSD words and constructions than other translators do. There are three arguments for this hypothesis. First, the subtitles are mostly intended for a Flemish audience only, so the need to comply with the strictest linguistic norms of the dominant variety in the Dutch language area might be smaller in Flemish subtitles compared to regular written translations that are distributed in a larger area. Second, because the original auditory signal remains present, subtitlers are aware that their

subtitles run a lower risk of being misunderstood (or not understood at all) compared to regular translations (where the original is replaced by the translations). Finally, the original colloquial footage might stimulate the subtitlers to retain a Belgian-Dutch colloquial word or construction that is used in the spoken source text.

➢ **Hypothesis 2:** the norm-adhering tendency in subtitles (cf. Hypothesis 1) will be weakest when spontaneous speech is subtitled (vs. voice-over speech) [Hypothesis 2a], in entertainment programs (vs. journalistic programs) [Hypothesis 2b], and when Belgian-Dutch speakers are subtitled (vs. speakers of Netherlandic Dutch and English) [Hypothesis 2c], as in these contexts subtitlers are more exposed to Belgian-Dutch colloquialisms, and are thus triggered more often to re-use these variants.

4.2 Variable selection

To investigate linguistic norm adherence in written translations and non-translations and to compare this to the linguistic choices made in interlingual and intralingual subtitles, we compiled a list of eight profiles, consisting of at least one BSD variant and one CBD alternative. As mentioned in Section 3.2.1, our profile set is mainly based on the list of Delaere et al. (2012). Before adding the profiles to our profile set, however, *Taaladvies* was consulted to verify the normative status of each variant[23], in the belief that that the variants that are currently labelled CBD were also labelled CBD in 2000-2005 (i.e. the period of our corpus data). This resulted in a final set of eight linguistic profiles, the distribution of which across the translation modes is displayed in Table 16. The data extraction and validation resulted in a final data set of 1,933 relevant instances (SoNaR: n = 887; DPC: n = 1,046)

Profile	Variants	Label	Non-translated written Dutch	Translated written Dutch < English	Translated written Dutch < French	Dutch Subtitles
1	*akkoord gaan met*	BSD	45	11	30	32
	akkoord zijn met	CBD	6	0	1	13

---

[23] At the time this dissertation is printed, the normative status of *akkoord zijn met* has changed into BSD.

2	part + aux + inf	BSD	29	16	19	91
	aux + inf + part	BSD	33	19	59	41
	aux + part + inf	CBD	27	3	3	20
3	*zulke* + plural noun	BSD	57	39	19	173
	zo'n + plural noun	CBD	3	0	0	11
4	*een beroep doen op*	BSD	87	29	108	31
	beroep doen op	CBD	18	3	7	6
5	*proberen te* + inf	BSD	19	9	15	117
	proberen + inf	CBD	2	1	2	1
6	*op het eerste gezicht*	BSD	13	10	4	32
	op het eerste zicht	CBD	3	0	0	5
7	*beginnen te* + inf	BSD	14	6	10	54
	beginnen + inf	CBD	9	2	4	26
8	*zodra*	BSD	93	37	103	225
	van zodra	CBD	13	1	5	9

Table 16. List of the profile frequencies per translation mode in case study 1

In what follows, a detailed description of the profiles will be given, by adding the translation and a corpus example for each variant. Van Dale dictionary (2015) was consulted for the translation and the example sentences were extracted from the SoNaR Corpus. For each example, a code is given that refers to the original corpus document from which the sentence was extracted.

Profile 1:	*akkoord gaan met – akkoord zijn met*
Translation:	*to agree with*
Corpus example BSD:	*Als er in de zaal zijn die niet **akkoord gaan met** mijn uiteenzetting, moeten ze het mij zeggen.*
	'If there is anyone in this room who doesn't **agree with** my statement, they should tell me.'
Document n°:	WR-P-E-G-0000003253
Corpus example CBD:	*Ik **ben** helemaal **akkoord met** het feit dat hij die genomen heeft.*
	'I totally **agree with** the fact that he has taken this one.'
Document n°:	WR-P-E-G-0000004909

Profile 2:	**part + aux + inf – aux + inf + part – aux + part + inf**
Translation:	position of the participle in the verbal end group
Corpus example BSD:	*Door zijn ziekte is de jonge Aleksej vaak zo zwak dat hij **gedragen moet worden.***
	'Because of his illness, the young Aleksej is so weak that he **has to be carried.**'
Document n°:	WR-P-E-G-0000006118
Corpus example BSD:	*Maar toch verwacht Patton dat hij binnen de kortste keren opzij **zal worden geschoven.***

'Yet Patton expected that he **will be pushed** aside in no time.'

Document n°:	WR-P-E-G-0000008460
Corpus example CBD:	*Het is een combinatie van plezier en iets dat **moet gedaan worden.***
	'It is a combination of fun and something that **should be done.**'
Document n°:	WR-P-E-G-0000000245

Profile 3:	***zulke* + meervoud – *zo'n* + meervoud**
Translation:	*such* + plural noun
Corpus example BSD:	*Maar waarom nemen ze **zulke risico's** om aan land te komen?*
	'But why do they take **such risks** to come on land?'
Document n°:	WR-P-E-G-0000003733
Corpus example CBD:	*Heb je zelf nooit **zo'n momenten** gehad dat je dacht: stel dat het wel waar zou zijn?*
	'Did you ever have had **such moments** when you thought, what if that would be true?'
Document n°:	WR-P-E-G-0000003789

Profile 4:	***een beroep doen op – beroep doen op***
Translation:	*to make (an) appeal to*
Corpus example BSD:	*Ze **deed een beroep op** de meest onverschrokken ontdekkingsreiziger uit die tijd.*
	'She **made an appeal to** the most intrepid explorer of that time.'
Document n°:	WR-P-E-G-0000007729
Corpus example CBD:	*Voor de scheepsinterieurs **deed** de CMB vaak **beroep op** architecten.*
	'For the ship's interiors, the CMB often **makes an appeal to** architects.'
Document n°:	WR-P-E-G-0000000243

Profile 5:	***proberen* + *te* + inf – *proberen* + inf**
Translation:	*to try (to)* + infinitive
Corpus example BSD:	*Ik ga het in de mond **proberen te** steken.*
	'I will **try to put** it into the mouth.'
Document n°:	WR-P-E-G-0000002265
Corpus example CBD:	*Wij hebben hier met de Spanjaarden een nieuw project **proberen opzetten.***
	'We **have tried to set up** here a new project with the Spaniards.'
Document n°:	WR-P-E-G-0000003318

Profile 6:	***op het eerste gezicht – op het eerste zicht***
Translation:	*at first sight*
Corpus example BSD:	***Op het eerste gezicht** lijkt het een woestijn, dor en levenloos.*
	'**At first sight** it looks like a desert, barren and lifeless.'
Document n°:	WR-P-E-G-0000003102

Corpus example	*Op het eerste zicht oogt hij normaal, maar zijn tenen zijn aan elkaar gegroeid, zijn oren zijn gespleten en zijn immuunsysteem is zwaar beschadigd.* '**At first sight** he looks normal, but his toes are grown together, his ears are split and his immune system is severely damaged.'
Document n°:	WR-P-E-G-0000006217

Profile 7:	***beginnen + te + inf – beginnen + inf***
Translation:	*to start* (*to*) + infinitive
Corpus example BSD:	*Het leger **begint te muiten.*** 'The army **starts to mutiny.**'
Document n°:	WR-P-E-G-0000000711
Corpus example CBD:	*Volgens de overheid zijn de gijzelnemers **beginnen schieten.*** 'According to the government, the hostage takers **started to shoot.**'
Document n°:	WR-P-E-G-0000006516

Profile 8:	***zodra – van zodra***
Translation:	as soon as
Corpus example BSD:	***Zodra** ze beseft dat er iets fout is gegaan, is ze terug.* '**As soon as** she realizes that something has gone wrong, she's back.'
Document n°:	WR-P-E-G-0000010103
Corpus example CBD:	***Van zodra** ze geleverd zijn, wordt de vennootschap opgedoekt.* '**As soon as** they are delivered, the company disappears.'
Document n°:	WR-P-E-G-0000005336

4.3   Results and discussion

In this section, the results of the profile-based correspondence analysis are presented and discussed. This statistical technique allows us to investigate whether subtitlers opt more often for CBD than for BSD, and whether these linguistic choices are comparable to the linguistic choices made in regular translations and non-translations. The two-dimensional plot will enable us to visually explore the linguistic distances between the relevant translation modes (subtitles, written translations, and written non-translations). The rationale behind this kind of plots is that the distance between the translation modes is smaller if the proportions of the chosen linguistic variants for each of the eight profiles in those translation modes are similar. The position of the translation modes relative to the position of the linguistic variants in the plot informs us of the linguistic options that are most often used in these contexts: the closer a translation mode is to certain variants,

the more often these variants are used in this translation mode in comparison with the other. In addition to providing some general information on the dispersion of the selected variants in relation to the position of the translation modes, Section 4.3.1 examines the first hypothesis, viz. that Flemish subtitlers are less norm-adhering than translators of regular written texts. In Sections 4.3.2 and 4.3.3, the second hypothesis is tested, viz. whether subtitles contain more CBD when spontaneous speech is subtitled (vs. voice-over speech), in entertainment programs (vs. journalisitc programs), and when Belgian-Dutch speakers are subtitled (vs. speakers of Netherlandic Dutch and English).

4.3.1   Linguistic norm adherence in different translation modes
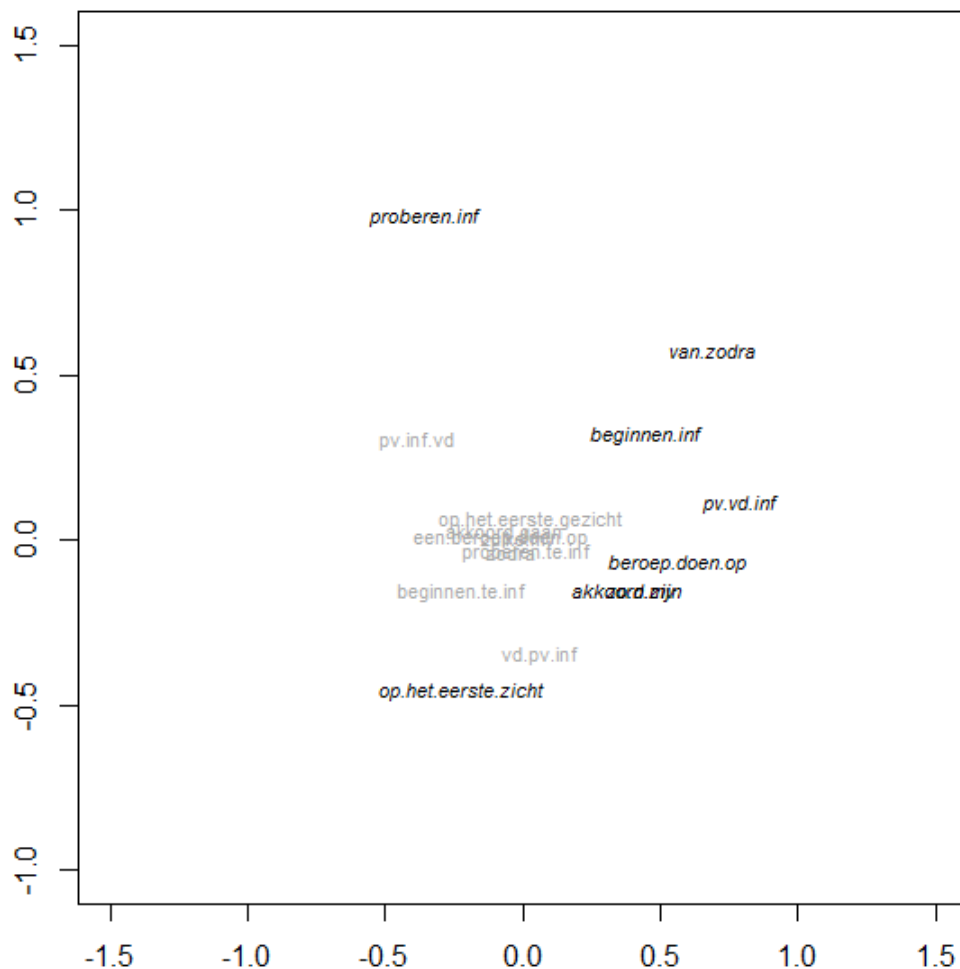
Figure 4. Distribution of the eight linguistic profiles in the SoNaR Corpus and the DPC (grey = BSD, black = CBD)

In Figure 4, the distribution of the linguistic variants is visualized[24]. The BSD variants are marked in grey, the CBD variants are marked in black. If we look at the dispersion of these linguistic items in the plot, we can see that the BSD variants are situated close to each other in the plot's origin, whereas the CBD variants are somewhat more widely distributed, but still mainly located in the right half of the plot, with both the items *proberen.inf* 'to try' and *op het eerste zicht* 'at first sight' as outliers. As correspondence analysis is basically a data reduction technique, which implies that some data is lost and the representation in only two dimensions is merely an estimation of the associations in the original data frame, we also have to evaluate the quality of this visualization. Therefore, we generated a so-called scree plot, which visualizes the degree of representativeness of the plot with respect to the total variation in the data set. The bars of the scree plot show how much of the total variation is associated with each dimension. As a consequence, the scree plot indicates how many dimensions are needed to reach a threshold (e.g. 80%).
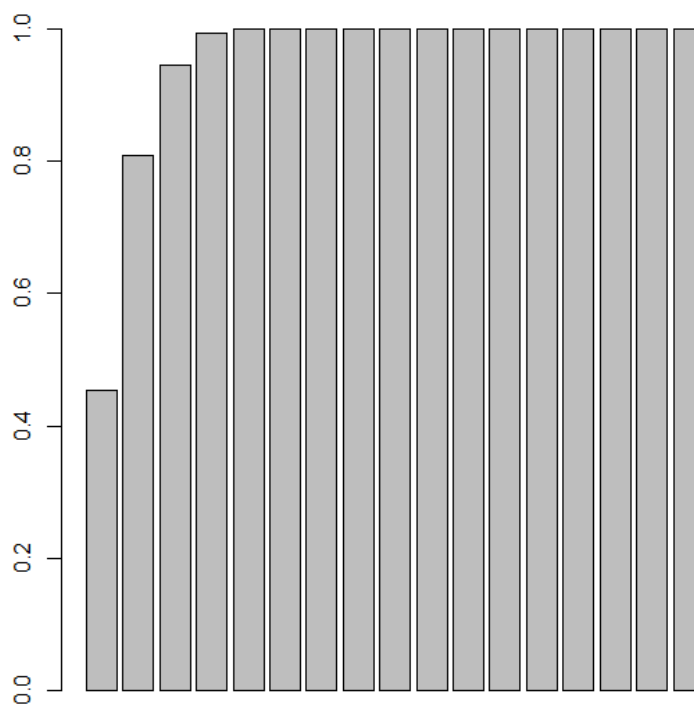
Figure 5. Cumulative scree plot for the distribution of the eight linguistic profiles in the SoNaR Corpus and the DPC

---

[24] All figures are also online available in color on http://www.eqtis.ugent.be/lynn_prieels/

The scree plot in Figure 5 demonstrates that only two dimensions are required to attain 80%, so we can reliably state that the two-dimensional plot in Figure 4 gives an accurate visualization of the original variation. The first dimension (x-axis) accounts for 45.42% of the variation, whereas the second axis (y-axis) captures 35.48%, which yields a total of 80.90%.
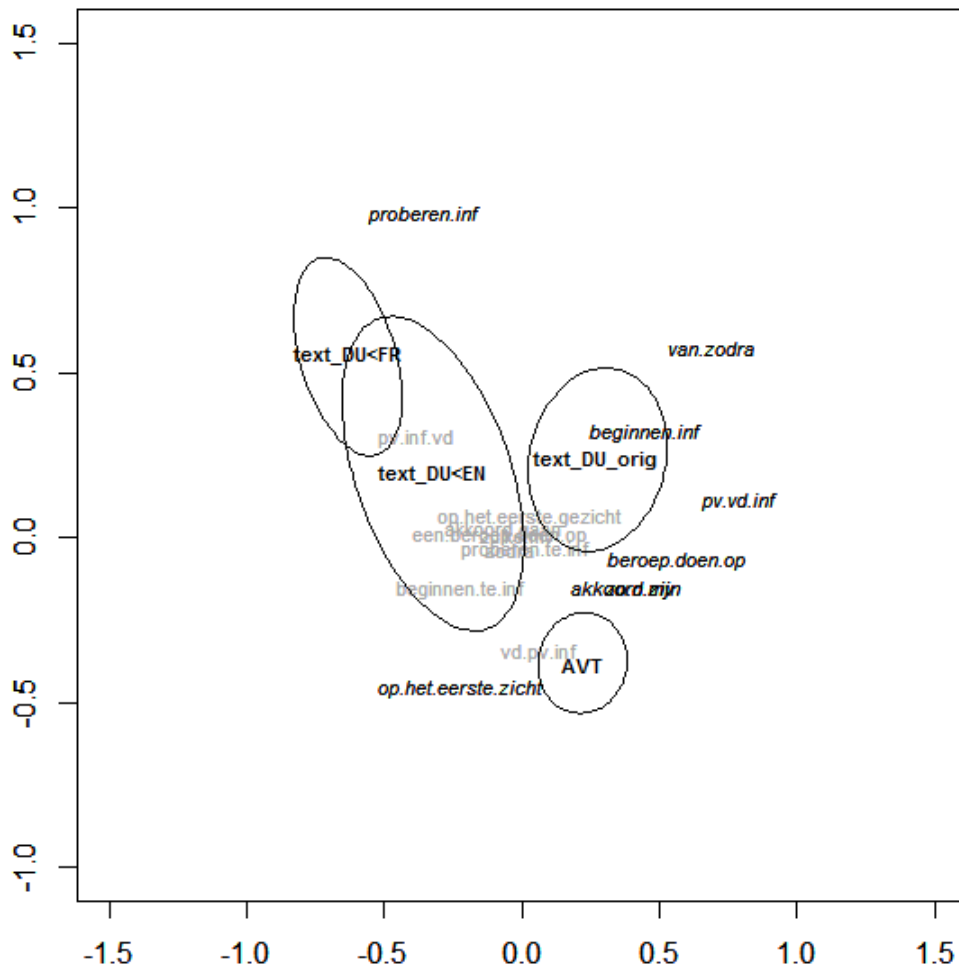
Figure 6. Biplot of the eight linguistic profiles and the four translation modes in the SoNaR Corpus and the DPC (grey = BSD, black = CBD)

In Figure 6, the different translation modes are plotted onto the individual variants, thus resulting in a biplot, which reveals two main findings. First, the linguistic choices made in translated texts (*text_DU<FR* and *text_DU<EN*), non-translated texts (*text_DU_orig*), and interlingual and intralingual subtitles (*AVT*) are significantly different (p < .05), as the ellipses of these translation modes do not overlap. There is, however, an overlap between the ellipses of Dutch texts translated from French (*text_DU<FR*) and Dutch texts translated

from English (*text_DU<EN*), which means that the linguistic choices of translators do not differ significantly according to the source language. Second, and more specifically, if we look at the position of the different translation modes (ellipses) relative to the position of the linguistic variants, it can be seen that both original (non-translated) Dutch texts (*text_DU_orig*) and subtitles (*AVT*) are clearly related to CBD variants is, as these ellipses are located the closest to the colloquial variants. Nevertheless, these translation modes are also surrounded by some of the BSD variants. Translations from English, on the other hand, are mostly related to BSD, the distance to the colloquial variants being larger (compared to subtitles and non-translations). The most norm-adhering text type in our data, although the difference with translations from English is not significant, are translations from French, as the distance from this variety to the CBD variants is the largest (and much larger compared to the distance to the BSD variants).

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

4.3.2 Contextual parameters influencing linguistic norm adherence in subtitles: main effects

Although the results in the previous section revealed that audiovisual translation (subtitles) in the Flemish context is less norm-adherent than other written translations, it is obvious that the analysis in the previous section is rather coarse-grained, as it does not take into account genre differences at all. In Delaere et al. (2012), the effect of genre on norm-adherent choices in Dutch translations and non-translations was analyzed, which showed for example that journalistic texts, irrespective of being translated or not,

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

As the effect of these parameters can only be measured for the subtitle part of our data set, we leave out the written translation data. Consequently, the positions of the linguistic variants and language varieties (ellipses) have to be recalculated for the subtitle data only. Figures 7 and 9 present the result of this new profile-based correspondence analysis.
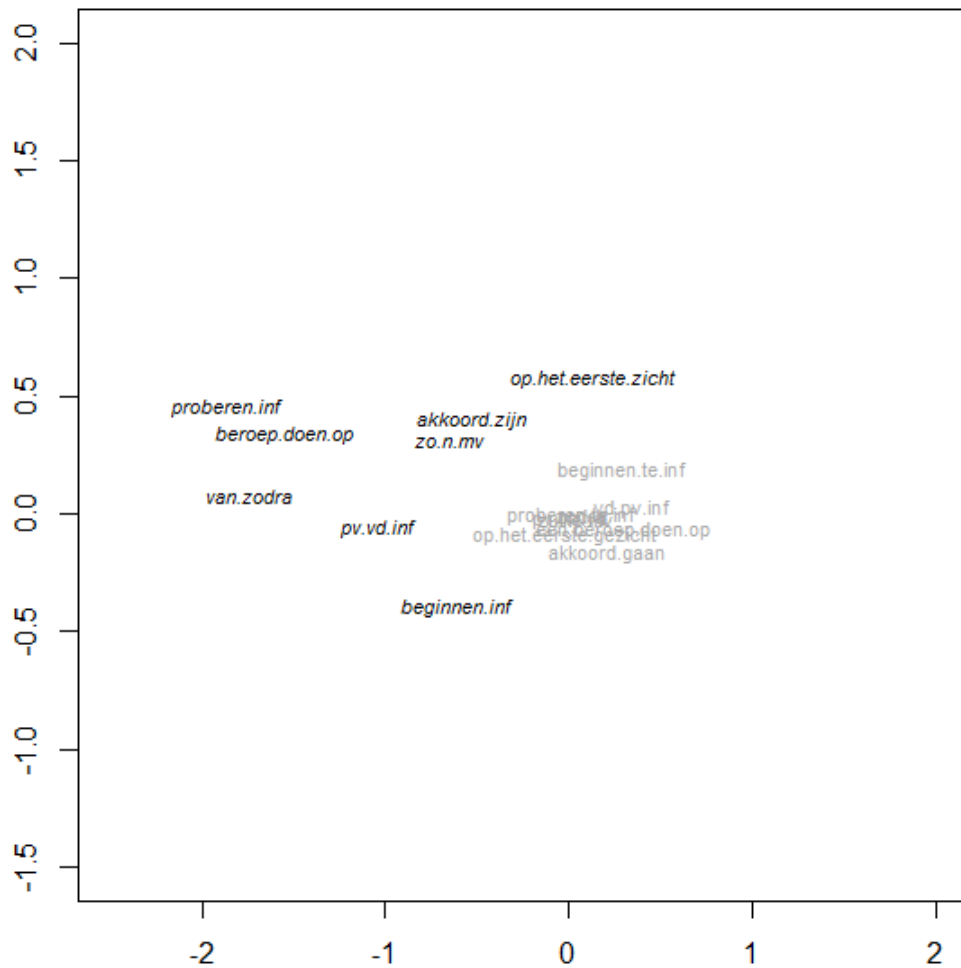
Figure 7. Distribution of the the eight linguistic profiles in the subtitle data of the SoNaR Corpus (grey = BSD, black = CBD)

Although the position of the variants indeed is somewhat different compared to Figures 4 and 6, the general picture remains more or less the same, with the BSD

variants located close to each other in the plot's origin, and the CBD variants more widely distributed mainly at the left-hand side of the plot.
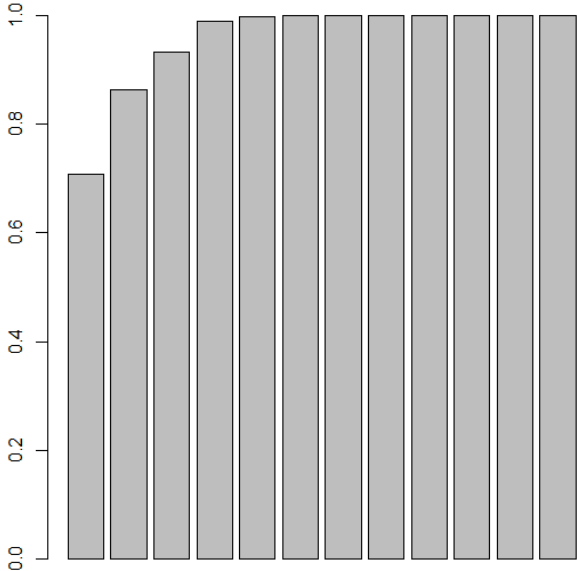
Figure 8. Cumulative scree plot for the distribution of the eight linguistic profiles in the subtitle data of the SoNaR Corpus

The scree plot in Figure 8 shows that the representativeness of the biplot in Figure 7 is very accurate. The visualization of merely two dimensions in the plot represents 86.47% of the total variation, with the x-axis accounting for 70.77% and the y-axis capturing 15.70%.
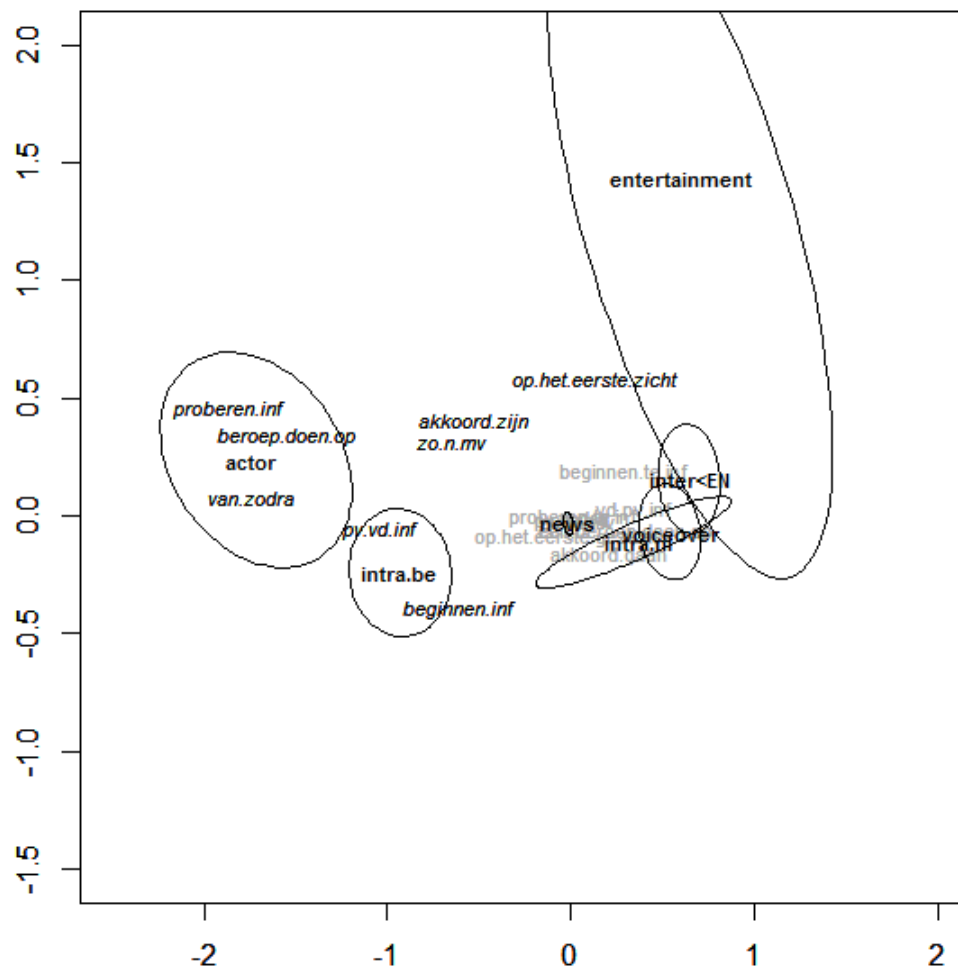
Figure 9. Biplot of the eight linguistic profiles and the contextual parameters *source language*, *speaker type*, and *program genre* in the subtitle data of the SoNaR Corpus (grey = BSD, black = CBD)

Figure 9 shows the position of the different context-specific subtitles relative to the position of the linguistic variants. It becomes immediately clear that subtitles of actors' speech (*actor*) are most clearly related to CBD variants, whereas subtitles of off-screen voice-over comments (*voiceover*) are related to BSD variants. Subtitles that are intralingual translations of Flemish speakers' speech (*intra.be*) are also most clearly related to CBD variants, albeit less outspoken than subtitles of actors' speech (as *intra.be* is closer to the BSD variants than *actor*). These two types of subtitles are thus less norm-adhering than the other types, which are much closer to the BSD variants: intralingual subtitles of Netherlandic-Dutch speakers (*intra.nl*), interlingual subtitles of English speech (*inter<EN*), and subtitles in news and entertainment programs (*news* and *entertainment*). Within this group of norm-adhering subtitling contexts, we can only see a significant

difference between subtitles in news programs and the other subtitling contexts, but there are no significant differences between all other subtitling contexts.

The lower degree of norm adherence in intralingual subtitles of Belgian-Dutch speakers as well as in subtitles of actors' speech can be explained by taking the nature of the original footage into account. First, actors' speech is, in contrast to voice-over speech, mostly spontaneous in nature (vs. monologic in voice-over speech), and thus has a greater chance of showing spontaneous, colloquial features. Furthermore, Van Hoof (2015) has repeatedly demonstrated that actors in Flemish fiction series speak a lot of tussentaal. From that perspective, it seems plausible to suggest that this causes the increased frequency of well-known and frequently attested CBD variants in the subtitles of Flemish actors' speech. Second, intralingual subtitles of Belgian-Dutch speakers have a greater chance of containing CBD variants than interlingual subtitles or intralingual subtitles of Netherlandic-Dutch speakers, as it is the only subtitling context in our data set where the subtitler is directly exposed to original Belgian-Dutch speech. As a consequence, it seems safe to conclude that the lower level of norm adherence in intralingual subtitles of Flemish speakers is caused by direct interference of the language use in the original Belgian-Dutch television program. However, we were not able to analyze the original source text, since the SoNaR Corpus does not contain the spoken television fragments. Therefore, we will set up a case study in which these explanations will be tested by using a new parallel corpus (cf. Chapter 6)

The significant differences between the subtitling contexts that are more closely related to the BSD variants are harder to interpret. It seems reasonable to state that the language choices made in the subtitles of news programs differ significantly from the language choices made in the other subtitling contexts because of the subtitlers' different linguistic preferences within the group of BSD variants. The ellipses of intralingual subtitles of Netherlandic-Dutch speech, the subtitles of voice-over speech, interlingual subtitles, and subtitles in entertainment programs are increasingly less related to the core of the BSD variants compared to the subtitles in news programs. Furthermore, especially the subtitles in entertainment programs seem to move away from the BSD core toward one of the CBD variants, viz. *op het eerste zicht* 'at first sight'. This suggests that all these subtitling contexts are norm-adhering to a large extent even though they are related to other linguistic variants.

In sum, the profile-based correspondence analysis of the subtitle data has confirmed that subtitles of actors' speech and Belgian-Dutch speech are less norm-adhering than subtitles of voice-over speech, and interlingual subtitles and intralingual subtitles of Netherlandic-Dutch speech respectively (verification of hypothesis 2a and 2c). On the other hand, it was only partly confirmed that subtitles in entertainment programs contain more colloquial variants than news programs. As mentioned, this is only true for one of the CBD variants (falsification of hypothesis 2b).

4.3.3 Contextual parameters influencing linguistic norm adherence in subtitles: interaction effects

The profile-based correspondence analysis in the previous section revealed interesting insights into norm-related linguistic choices made in certain specific subtitling contexts. What the analysis in Section 4.3.2 did not reveal, however, are the mutual relationships or interactions between these different subtitling contexts with respect to linguistic norm adherence. What we do not know yet, for instance, is (i) whether Belgian actors are subtitled differently compared to Belgian voice-overs and (ii) whether voice-overs in entertainment programs are subtitled differently compared to voice-overs in news programs.

To answer these questions, we computed two-way interactions between *source language* and *speaker type*, between *source language* and *genre* and between *genre* and *speaker type*, and visualized these interactions in three biplots (Figures 10, 11, and 12). The position of the linguistic variants in these interaction plots remains unchanged compared to the main plot in Figure 7.
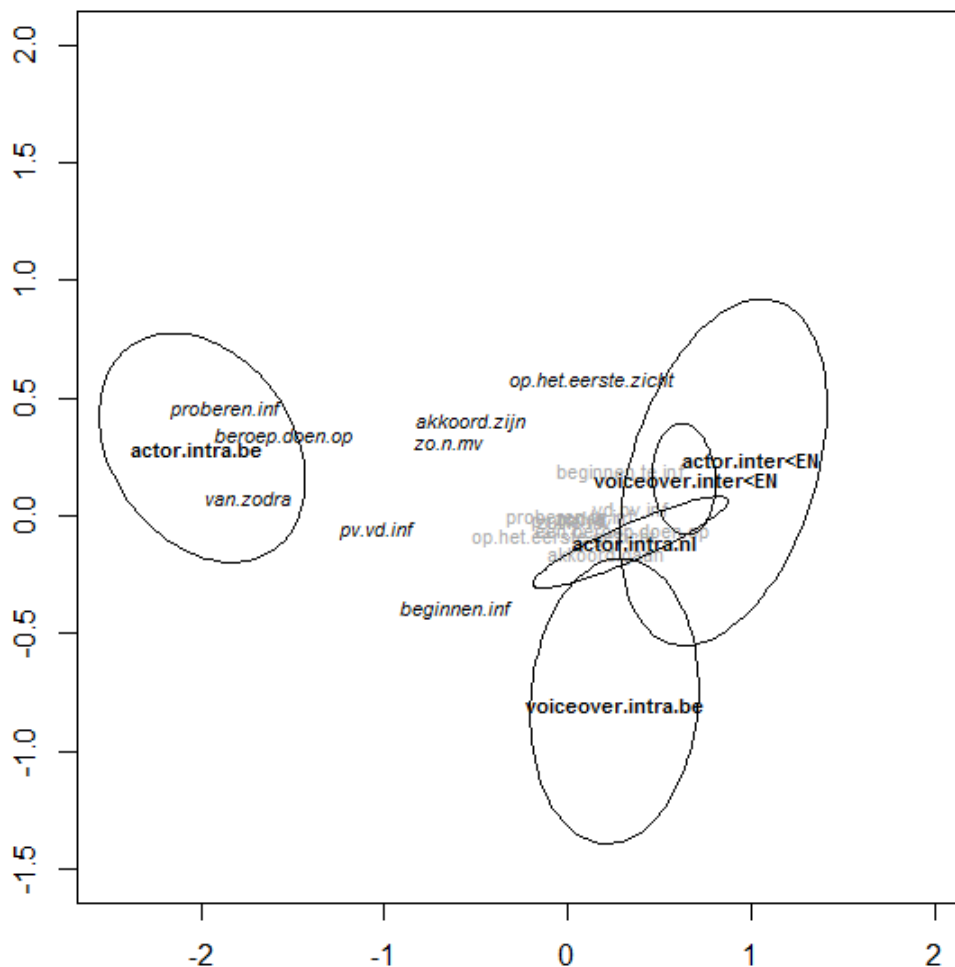
Figure 10. Biplot of the eight linguistic profiles and the interaction between *speaker type* and *source language* in the subtitle data of the SoNaR Corpus (grey = BSD, black = CBD)

Figure 10 shows the interaction effect between the *speaker type* and *source language*. The most interesting observation is that intralingual subtitles of Belgian (or Flemish) actors' speech (*actor.intra.be*) are clearly related to CBD variants, whereas intralingual subtitles of Belgian voice-over speech (*voiceover.intra.be*) are located much closer to the BSD variants. In other words, subtitles contain more colloquial variants if the original speech is delivered by a Belgian-Dutch speaker in a dialogic (colloquial) context than in a monologic context. BSD variants are most frequently attested in subtitles of Netherlandic actors' speech (*actor.intra.nl*). Interlingual subtitles of voice-over speech (*voiceover.inter<EN*) and interlingual subtitles of actors' speech (*actor.inter<EN*) are somewhat less norm-adherent, as these subtitling contexts are more closely related to one of the CBD variants (*op het eerste zicht* 'at first sight'). Within this group of norm-

adhering subtitling contexts, only the ellipses of intralingual subtitles of Belgian-Dutch voice-over speech and interlingual subtitles of voice-over speech do not overlap, which implies that subtitlers prefer other BSD variants when subtitling a Flemish voice-over compared to an English voice-over. There are no significant differences between the other subtitling contexts.

The main conclusion emerging from Figure 10, building on the relative distances between the ellipses, is that subtitles of Belgian-Dutch actors' speech is the only subtitle type that is clearly characterized by CBD variants. We see two potential explanations for this, one in terms of interference or shining-through, one in terms of normalization. The first explanation would be that subtitlers transfer the colloquial variants from the spoken material into the subtitles, simply because this linguistic material is available in the original footage. This obviously can only be the case if it can be shown that the original speech of the Belgian-Dutch actors contains more CBD variants than all other types of speech (including Belgian-Dutch voice-over speech). In Chapter 6, we will test this assumption. The second explanation would be that subtitlers strategically transfer these CBD variants to create the spontaneous, colloquial style which is typical for dialogic contexts, and replace the colloquial variants by BSD variants in monologic contexts. This can only be the case if it can be shown that all Belgian-Dutch source material (actors and voice overs) contains an equal amount of CBD variants. As this case study has not analyzed the original speech in detail, it is impossible to say which of the explanations is most plausible. Nevertheless, one could reasonably argue that the second explanation is less plausible (viz. translators use CBD variants strategically), given the situation in Figure 10, as we would then expect that all subtitles of actors' speech, irrespective of the source language, would contain many more CBD variants than subtitles of voice-over speech – quod non. In case study 3, the linguistic choices made in the subtitles will be compared to the original spoken source text in order to verify whether the first explanation is indeed most accurate.
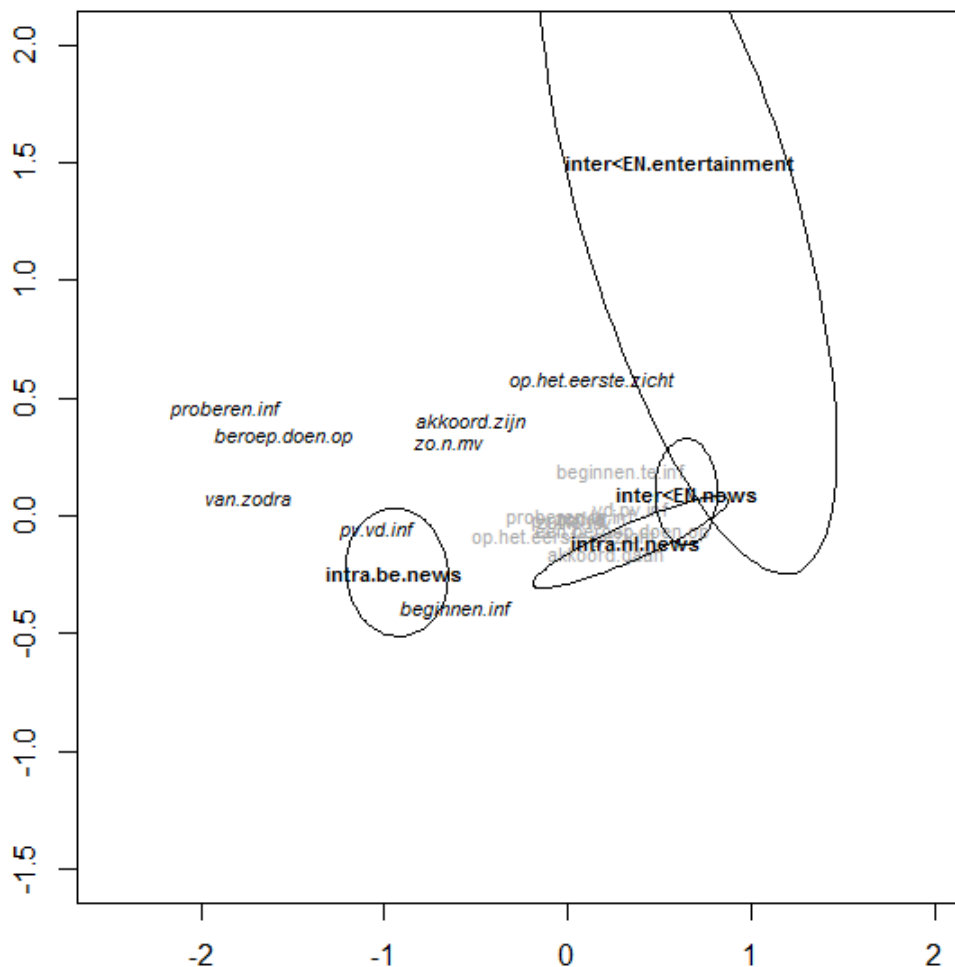
Figure 11. Biplot of the eight linguistic profiles and the interaction between *source language* and *program genre* in the subtitle data of the SoNaR Corpus (grey = BSD, black = CBD)

Figure 11 shows the interaction between *source language* and *program genre*. There is only a significant difference between the linguistic choices in subtitles of Belgian speakers in news programs (*intra.be.news*) and all other subtitle types (*intra.nl.news, inter<EN.entertainment, inter<EN.news*). News programs in which Belgian-Dutch speakers are subtitled are related most to the CBD variants while interlingual and intralingual Netherlandic subtitles of news programs and interlingual subtitles of entertainment programs are located much closer to the BSD variants. This supports the interference explanation mentioned above, and suggests that subtitlers do not strategically add colloquial variants in subtitles when the original footage does not contain CBD variants at all.
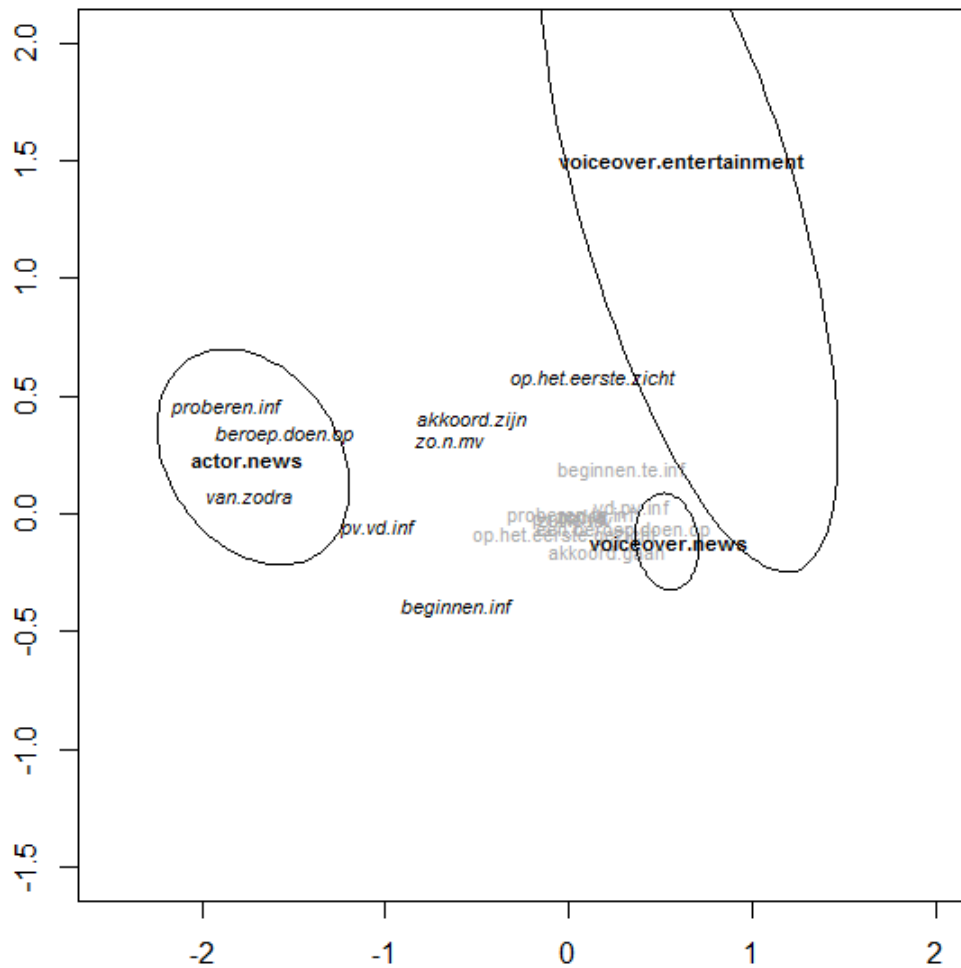
Figure 12. Biplot of the eight linguistic profiles and the interaction between *speaker type* and *program genre* in the subtitle data of the SoNaR Corpus (grey = BSD, black = CBD)

Finally, Figure 12 presents the interaction between *speaker type* and *program genre*, in which it can be observed that subtitles of interviewees' speech in news programs (*actor.news*) contain much more CBD variants than voice-overs' speech in news (*voiceover.news*) and entertainment programs (*voiceover.entertainment*), leading to the conclusion that CBD variants most frequently show up in subtitles of spontaneous, dialogic speech.

4.4   Concluding remarks

Building on two large corpora of Belgian-Dutch written and audiovisual translation, this case study analyzed how subtitlers, translators and original authors deal with norm-related language variation in Flanders. By investigating the dispersion of BSD vs. CBD in subtitling and in other written translations and non-translations, it is demonstrated which text types conform to the BSD norm and which exhibit more Belgian-Dutch colloquialisms. The hypothesis that translators are more norm-adherent than non-translators was verified, but it was also shown that there is a considerable difference in linguistic use between audiovisual and written translation. Subtitle data contained significantly more CBD variants compared to regular written translations. In-depth analyses pointed out that linguistic choices in these subtitles are mainly determined by the source language and by the speaker type. If the source language of the original footage is Belgian Dutch (yielding an intralingual translation), the amount of colloquial variants increased significantly (compared to interlingual translations from English and intralingual translations from Netherlandic Dutch). Additionally, if the subtitled voice is of an actor or interviewee, the frequency of CBD variants also increased significantly. The most plausible explanation offered for these results is that subtitlers (consciously or unconsciously) transfer the colloquial variants in the original footage directly to the subtitles, thereby maintaining the 'Belgian atmosphere' in the original footage. However, since the SoNaR Corpus does not contain the spoken television fragments, we were not able to analyze the original source text. In Chapter 6, we will therefore build a new parallel corpus and set up a case study in which the language use in the spoken source text will be compared to the corresponding subtitles in order to verify the aforementioned assumptions.

Chapter 5
Case study 2: comparing lexical and grammatical norm adherence in Belgian-Dutch subtitles

The results of the first case study revealed that Flemish subtitlers use more Colloquial Belgian Dutch (CBD) than translators of other written text genres. We do not know, however, whether the use of colloquial language in subtitling is mainly situated on a lexical level, on a grammatical level, or on both. Furthermore, case study 1 is based on a small subcomponent of the subtitle corpus in SoNaR (*n* = 2,048,480), since the web tool OpenSoNaR was not yet publicly available at that moment. As a result, we applied a rather general genre classification in which we dived the television programs into *journalism* or *entertainment.* For this second case study, however, we consulted the entire subtitle component of the SoNaR Corpus (*n* = 18,687,891) and we divided the television programs into five specific genres that are largely based on the genre classification of Creeber (2008). This second case study allows us not only to investigate the extent to which Flemish subtitlers use CBD variants instead of Belgian Standard Dutch (BSD) variants, but also whether the subtitles contain more colloquial lexemes than colloquial grammatical constructions. Since VRT's subtitling guidelines tolerate the use of colloquial lexicon in certain contexts, whereas colloquial grammatical constructions must always be converted into standard language, this study examines the extent to which Flemish subtitlers adhere to this policy. In addition, it investigates which contextual factors (e.g. *source language* and *program genre*) have an influence on the subtitlers' lexical and grammatical choices.

5.1 Hypotheses

Against the background of the subtitling practice on Flemish television described in Chapter 2 and on the basis of the results that were obtained in the first case study, three hypotheses are put forward:

> **Hypothesis 1:** subtitles on Flemish television contain fewer CBD lexemes than CBD grammatical constructions. Although VRT's subtitling guidelines allow for the reproduction of colloquial lexemes in the subtitles to some extent, they suggest that the sensitivity toward the use of standard and colloquial lexemes is remarkably higher than toward the use of standard and colloquial constructions. The use of colloquial lexicon is merely 'more or less' tolerated and 'editorial agreement is required before the use a colloquial lexeme is allowed' (Dewulf & Saerens 2000). As a consequence, it can be expected that the use of colloquial lexical variants is restricted to a minimum, as the subtitlers do not want to run the risk of making mistakes. This sensitivity does not apply to colloquial grammatical constructions, since the subtitling guidelines merely mention that 'morphosyntactic colloquialisms are always translated into standard language'. Furthermore, lexical features are more salient than grammatical features (Lybaert 2014b), which means that subtitlers easily detect these colloquial lexemes, whereas colloquial grammatical constructions are easily missed by the subtitlers because their awareness toward colloquial grammar is less strong. This lower sensitivity toward grammatical colloquialisms in VRT's subtitling guidelines and the lower salience of grammatical features in general, makes us to assume that these features will be reproduced more often in the subtitles.

> **Hypothesis 2:** the use of standard and colloquial grammatical constructions is less context-dependent than the use of standard and colloquial lexemes. Besides the assumption that grammatical colloquialisms occur more frequently in Flemish subtitles than lexical colloquialisms, it can also be expected that these CBD grammatical items are less tied to specific contexts. Lybaert (2014b) showed that syntactical features are less salient than lexical features are, which can be attributed to the abstract nature of syntactic elements (Van Bree 2000): unlike the lexicon, this domain is characterized by abstract rules, which makes it more automated or less concrete, so that language users unconsciously use and perceive these grammatical constructions. As a consequence, these grammatical features frequently occur in various situations in daily language use without being perceived very consciously; therefore, it can be expected that these elements are also frequently reproduced in different subtitling contexts.

> ➢ **Hypothesis 3:** the number of CBD variants in the subtitles will be higher in entertainment, infotainment and humor programs (vs. informative programs) [Hypothesis 3a], in programs with a general audience (vs. children) [Hypothesis 3b], when the spontaneous speech of non-actors (vs. scripted language of actors) is subtitled [Hypothesis 3c], and when Belgian-Dutch speakers (vs. speakers of Netherlandic Dutch and English) are subtitled [Hypothesis 3d]. On the one hand, the public broadcaster VRT allows the use of non-standard varieties in fiction series and other entertainment programs, whereas standard language is required in informative programs and in TV programs intended for children (Hendrickx 1998). On the other hand, subtitlers are more likely to be exposed to colloquial variants in the aforementioned contexts, which increases the odds that they reuse these variants in the subtitles (cf. case study 1).

5.2   Variable selection

In order to investigate the use of BSD vs. CBD in the subtitles, we compiled three sets of linguistic profiles. The tables below provide an overview of the lexical-paradigmatic profiles (Table 17),   the constructional-paradigmatic profiles (Table 18), and the syntagmatic profiles (Table 19) that were used for this case study. All variants were extracted from the SoNaR Corpus, which resulted in a final data set of 36,551 validated observations (BSD: *n* = 27,142; CBD: *n* = 9,409). Tables 17 to 19 provide an overview of the number of attestations of each variant. In addition, a detailed description of the profiles is given underneath each table. The example sentences were extracted from the SoNaR Corpus and for each example, a code is given that refers to the original corpus document from which the sentence was extracted.

Profile	Belgian Standard Dutch	Colloquial Belgian Dutch	Translation or meaning
1	*autosnelweg* (*n* = 33) *autoweg* (*n* = 90) *snelweg* (*n* = 129)	*autostrade* (*n* = 34)	*motorway*
2	*bestelwagen* (*n* = 99) *bestelauto* (*n* = 3)	*camionette* (*n* = 234)	*delivery van*
3	*fiets* (*n* = 1257) *rijwiel* (*n* = 7)	*velo* (*n* = 61)	*bicycle*
4	*handtas* (*n* = 186)	*sacoche* (*n* = 157)	*handbag*
5	*jas* (*n* = 643)	*frak* (*n* = 51)	*coat*

6	*krant* (*n* = 1026) *dagblad* (*n* = 20)	*gazet* (*n* = 232)	*newspaper*
7	*laars* (*n* = 90)	*bot* (*n* = 24)	*boot*
8	*motor* (*n* = 392) *motorfiets* (*n* = 43)	*moto* (*n* = 206)	*motorbike*
9	*nieuwsgierig* (*n* = 320) *benieuwd* (*n* = 767)	*curieus* (*n* = 142)	*curious*
10	*oom* (*n* = 334) *ome* (*n* = 8)	*nonkel* (*n* = 634)	*uncle*
11	*het platteland* (*n* = 147)	*de buiten* (*n* = 44)	*countryside*
12	*schrikken* (*n* = 894)	*verschieten* (*n* = 442)	*to be frightened*
13	*stropdas* (*n* = 12) *das* (*n* = 148)	*plastron* (*n* = 31)	*tie*
14	*vrachtwagen* (*n* = 412)	*camion* (*n* = 127)	*truck*
15	*wastafel* (*n* = 14) *wasbak* (*n* = 11)	*lavabo* (*n* = 49)	*sink*

Table 17. List of the lexical-paradigmatic profiles used in case study 2

Profile 1:	*autosnelweg – autoweg – snelweg - autostrade*
Translation:	*motorway*
Corpus example BSD:	*Op de **autosnelweg** ligt de snelheid hoger.* 'The speed is higher on the **motorway**.'
Document n°:	WR-P-E-G-0000006729
Corpus example BSD:	*We sluiten de **autoweg** af in laatste instantie.* 'As a last resort, we block the **motorway**.'
Document n°:	WR-P-E-G-0000008884
Corpus example BSD:	*Het verkeer op de **snelweg** rijdt vlot.* 'The traffic on the **motorway** is running smoothly.'
Document n°:	WR-P-E-G-0000004244
Corpus example CBD:	*Fietsen op een **autostrade** kan gewoon niet.* 'Bikes are not allowed on the **motorway**.'
Document n°:	WR-P-E-G-0000002048

Profile 2:	*bestelwagen – bestelauto – camionette*
Translation:	*delivery van*
Corpus example BSD:	*Al zijn persoonlijke spullen worden in een **bestelwagen** geladen.* 'All his personal belongings are loaded into a **delivery van**.'
Document n°:	WR-P-E-G-0000006255
Corpus example BSD:	*Maak de **bestelauto** open!* 'Open the **delivery van**!'
Document n°:	WR-P-E-G-0000010481

Corpus example CBD:	*Heeft hij donderdagavond met die **camionette** gereden?*
	'Did he drive the delivery van on Thursday night?'
Document n°:	WR-P-E-G-0000004543

Profile 3:   *fiets – rijwiel - velo*

Translation:	*bycicle*
Corpus example BSD:	*Ik reed met de **fiets** naar school.*
	'I rode to school by **bicycle**.'
Document n°:	WR-P-E-G-0000006836
Corpus example BSD:	*Bij dag en dauw halen fietsers hun **rijwiel** boven.*
	'Cyclists get on their bicycles before daybreak.'
Document n°:	WR-P-E-G-0000001921
Corpus example CBD:	*Mijn **velo** is gestolen!*
	'My **bicycle** was stolen!'
Document n°:	WR-P-E-G-0000006422

Profile 4:   *handtas - sacoche*

Translation:	*handbag*
Corpus example BSD:	*Mijn bankkaart zat in mijn **handtas**.*
	'My credit card was in my **handbag**.
Document n°:	WR-P-E-G-0000003202
Corpus example CBD:	*Ze heeft jou tenminste niet geslagen met haar **sacoche**.*
	'At least she did not beat you with her **handbag**.'
Document n°:	WR-P-E-G-0000000169

Profile 5:   *jas - frak*

Translation:	*coat*
Corpus example BSD:	*Hij droeg een das en een lederen **jas**.*
	'He was wearing a tie and a leather **coat**.'
Document n°:	WR-P-E-G-0000003771
Corpus example CBD:	*Ik kan mijn **frak** niet uitdoen.*
	'I can't take off my **coat**.'
Document n°:	WR-P-E-G-0000007256

Profile 6:   *krant – dagblad - gazet*

Translation:	*newspaper*
Corpus example BSD:	*'s Anderendaags stond dat in de **krant**.*
	'The next day, it was in the **newspaper**.'
Document n°:	WR-P-E-G-0000008663
Corpus example BSD:	*Ooit schreef een Nederlandstalig **dagblad** dat hij een arrogante Waal was.*
	'Once, a Dutch **newspaper** wrote that he was an arrogant Walloon.
Document n°:	WR-P-E-G-0000002659

Corpus example CBD:	*Laat me mijn **gazet** lezen.*
	'Let me read my **newspaper**.'
Document n°:	WR-P-E-G-0000005592

Profile 7: *laars - bot*

Translation:	*boot*
Corpus example BSD:	*We hebben die **laars** gevonden op het strand.*
	'We found that **boot** on the beach.'
Document n°:	WR-P-E-G-0000007287
Corpus example CBD:	*De Hollandse koningin trekt haar caoutchouc **botten** aan en gaat de mensen helpen.*
	'The Dutch queen puts on her rubber **boots** and is going to help the people.'
Document n°:	WR-P-E-G-0000000722

Profile 8: *motor – motorfiets - moto*

Translation:	*motorbike*
Corpus example BSD:	*Ze krijgen een dolle tocht achter op de motor.*
	'They will have a wild ride on the motorbike.'
Document n°:	WR-P-E-G-0000002324
Corpus example BSD:	*Vanaf je 18 mag je rijden met een **motorfiets** met beperkt vermogen.*
	'At the age of 18 you can ride a **motorbike** with limited power.'
Document n°:	WR-P-E-G-0000010145
Corpus example CBD:	*Ze kregen de **moto** niet meer aan de praat.*
	'They could not get the **motorbike** to start.'
Document n°:	WR-P-E-G-0000003771

Profile 9: *nieuwsgierig – benieuwd – curieus*

Translation:	*curious*
Corpus example BSD:	*Ik was **nieuwsgierig** naar de resultaten.*
	'I was **curious** about the results.'
Document n°:	WR-P-E-G-0000008629
Corpus example BSD:	*Ik ben echt **benieuwd** naar jouw indrukken.*
	'I am really **curious** about your impressions.'
Document n°:	WR-P-E-G-0000007182
Corpus example CBD:	*Je maakt me **curieus**.*
	'You are making me **curious**.'
Document n°:	WR-P-E-G-0000002829

Profile 10: *oom – ome – nonkel*

Translation	*uncle*
Corpus example BSD:	*Kim en Gemma hebben hun **oom** Dell in jaren niet gezien.*
	'Kim and Gemma haven't seen their **Uncle** Dell for years.

Document n°:	WR-P-E-G-0000005344
Corpus example BSD:	*Hij heeft de neus van **ome** Gerrit.*
	'He has **Uncle** Gerrit's nose.'
Document n°:	WR-P-E-G-0000002817
Corpus example CBD:	*Tante Ria en **nonkel** Bob waren beste vrienden.*
	'Aunt Ria and **Uncle** Bob were best friends.'
Document n°	WR-P-E-G-0000004660

Profile 11:	*het platteland – de buiten*
Translation:	*the countryside*
Corpus example BSD:	*Ik woon in de stad , maar wil graag op **het platteland** wonen.*
	'I live in the city, but I want to live in **the countryside**.'
Document n°:	WR-P-E-G-0000003884
Corpus example CBD:	*Je kan hier echt wandelen en het gevoel hebben dat je hier op **de buiten** bent.*
	'Here you can really walk and have the feeling that you are in **the countryside**.'
Document n°:	WR-P-E-G-0000000746

Profile 12:	schrikken – verschieten
Translation:	*to be frightened*
Corpus example BSD:	*Hij is nu vast zo **geschrokken** dat hij niet meer terugkomt.*
	'Now he is definitely so **frightened** that he will not come back.'
Document n°:	WR-P-E-G-0000011280
Corpus example CBD:	*Ik ben **verschoten** van wat ik gezien heb.*
	'I am **frightened** by what I have seen.'
Document n°:	WR-P-E-G-0000003816

Profile 13:	*stropdas – das – plastron*
Translation:	*tie*
Corpus example BSD:	*Een jasje en een **stropdas** zijn verplicht.*
	'A jacket and **tie** are obligatory.'
Document n°:	WR-P-E-G-0000010921
Corpus example BSD:	*Ze hebben zelfs mijn **das** en riem afgenomen!*
	'They even took my **tie** and belt!'
Document n°:	WR-P-E-G-0000002357
Corpus example CBD:	*Op 21 maart wil ik een groene **plastron** dragen.*
	'On March 21, I want to wear a green **tie**.'
Document n°:	WR-P-E-G-0000005411

Profile 14:	*vrachtwagen – camion*
Translation:	*truck*

Corpus example BSD:	*Deze **vrachtwagen** kwam ongehinderd door de grenscontrole.*
	'This **truck** made it through border control without hindrance.'
Document n°:	WR-P-E-G-0000004208
Corpus example CBD:	*Tot de grens zitten jullie in één **camion**, daarna stappen jullie over.*
	'You will ride in one **truck** to the border, then you will change vehicles.'
Document n°:	WR-P-E-G-0000002360

Profile 15:	*wastafel – wasbak – lavabo*
Translation:	*sink*
Corpus example BSD:	*Er staan geen potjes op de **wastafel** , dus hij is ordelijk.*
	'There are no pots in the **sink**, so it is tidy.'
Document n°:	WR-P-E-G-0000006068
	*Hij vond zijn kleren in de **wasbak**.*
Corpus example BSD:	'He found his clothes in the **sink**.'
	WR-P-E-G-0000008540
Document n°:	*Het warm water van de **lavabo** werkt niet.*
Corpus example CBD:	'The hot water from the **sink** does not work.'
Document n°:	WR-P-E-G-0000008741

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

Table 18. List of the constructional-paradigmatic profiles used in case study 2

Profile 1:	**adj + *om* + *te* + inf – adj + *om* + inf**
Translation:	adj + *to* + infinitive
Corpus example BSD:	*Kieling vindt het **ontroerend om te zien** hoe snel de dieren hem vertrouwen.*
	'Kieling finds it **touching to see** how quickly the animals trust him.'
Document n°:	WR-P-E-G-0000009945
Corpus example CBD:	*Dat is **goed om weten**.*
	'That is **good to know**.'
Document n°:	WR-P-E-G-0000004251

Profile 2:	**beginnen + te + inf – beginnen + inf**
Translation:	*to start* (+ *to*) + infinitive
Corpus example BSD:	*Ik ben ook vroeg **beginnen te werken**.*
	'I also **started to work** early.'
Document n°:	WR-P-E-G-0000007892

Corpus example CBD:	*Een deel van La Palma was **beginnen schuiven.***
	'A part of La Palma had **started to slide**.'
Document n°:	WR-P-E-G-0000007415

Profile 3: *een beroep doen op – beroep doen op*

Translation:	*to make (an) appeal to*
Corpus example BSD:	*Om die patiënten te vinden **doet** hij **een beroep op** dokter Rob Medaer.*
	'To find these patients, he **made an appeal to** doctor Rob Medaer.'
Document n°:	WR-P-E-G-0000007278
Corpus example CBD:	*Ze **doen** al jarenlang **beroep op** dezelfde figuranten.*
	'For many years now, they **have made an appeal to** the same extras.'
Document n°:	WR-P-E-G-0000008399

Profile 4: *durven + te + inf – durven + inf*

Translation:	*to dare (+ to) + infinitive*
Corpus example BSD:	*Niemand **durfde te stoppen** met applaudisseren.*
	'Nobody **dared to stop** applauding.'
Document n°:	WR-P-E-G-0000000645
Corpus example CBD:	*Zou jij dat **durven zeggen** op televisie?*
	'Would you **dare to say** that on television?
Document n°:	WR-P-E-G-0000007141

Profile 5: *mocht(en) – als – moest(en)*

Translation:	hypothetical clause
Corpus example BSD:	***Mocht** er iets mislopen , zorg dan goed voor mijn zoon Gawein.*
	'**If** anything goes wrong, please take care of my son Gawein.'
Document n°:	WR-P-E-G-0000005679
Corpus example BSD:	***Als** hij bleef , zouden de geallieerden hem arresteren of partizanen hem ombrengen.*
	'**If** he stayed, the allies would arrest him or the partisans would kill him.'
Document n°:	WR-P-E-G-0000005194
Corpus example CBD:	***Moest** het nodig zijn , dan kan een speciale scan gemaakt worden.*
	'**If** it is necessary, a special scan can be made.'
Document n°:	WR-P-E-G-0000002010

Profile 6: *niet hoeven – niet moeten*

Translation:	*not have to*
Corpus example BSD:	*Je **hoeft niet** veel te bezitten om al een doelwit te zijn.*
	'You **do not have to** possess a lot to be a target.'
Document n°:	WR-P-E-G-0000002225

Corpus example	*De kinderen **moeten niet** meer bij mama wonen.*
	'The children **do not have to** live with their mother anymore.'
Document n°:	WR-P-E-G-0000001062

Profile 7: *op het eerste gezicht – op het eerste zicht*

Translation:	*at first sight*
Corpus example BSD:	***Op het eerste gezicht** lijkt niemand gewond.*
	'**At first sight** nobody looks injured.'
Document n°:	WR-P-E-G-0000006210
Corpus example CBD:	***Op het eerste zicht** zijn Slovaken vrij stuurs.*
	'**At first sight** Slovaks are quite surly.'
Document n°:	WR-P-E-G-0000005958

Profile 8: *over - na*

Translation:	*past* (time indication)
Corpus example BSD:	*Ik ben dan zoals gewoonlijk rond kwart **over** zeven naar mijn appartement gegaan.*
	'As usual, I went to my apartment at quarter **past** seven.
Document n°:	WR-P-E-G-0000003798
Corpus example CBD:	*Ontbijt om kwart **na** zes is vroeg.*
	'Breakfast at quarter **past** six is early.'
Document n°:	WR-P-E-G-0000007419

Profile 9: *passief – passief + geweest/geworden*

Translation:	passive clause
Corpus example BSD:	*Ook dit museum **is gebouwd** in opdracht van Leopold.*
	'This museum **was** also **built** under the authority of Leopold.'
Document n°:	WR-P-E-G-0000008926
Corpus example CBD:	*Het huisje **is** ooit **gekraakt geweest**.*
	'The cottage **was** once **broken** into by squatters.'
Document n°:	WR-P-E-G-0000005308

Profile 10: *vz + vz + en – vz + vz*

Translation:	preposition + preposition (+*en*)
Corpus example BSD:	*Hun tanden in het gehemelte zijn **naar achteren** gekromd.*
	'The teeth in their palate are bent toward the back.'
Document n°:	WR-P-E-G-0000002398
Corpus example CBD:	*Uit de ideeën van alle medewerkers zijn veel nieuwe initiatieven en projecten **naar voor** gekomen.*
	'A lot of new initiatives and projects have arisen from the employee's ideas.'
Document n°:	WR-P-E-G-0000006157

Profile 11:	*proberen + te + inf – proberen + inf*
Translation:	*to try (+ to) + infinitive*
Corpus example BSD:	*Hij zal **proberen te achterhalen** of tijgerhaaien kustbewoners zijn of in een ruimer gebied leven.*
	'He will **try to discover** whether tiger sharks live at the coast or in a larger area.'
Document n°:	WR-P-E-G-0000000268
Corpus example CBD:	*Door die zwangerschap van Leontien heb ik dat **proberen verdringen**.*
	'Because of Leontien's pregnancy, I **tried to suppress** that.'
Document n°:	WR-P-E-G-0000003044

Profile 12:	*zeker weten dat – er zeker van zijn dat – zeker zijn dat*
Translation:	*to be sure that*
Corpus example BSD:	*Jean **weet zeker dat** het om een groot dier gaat.*
	'Jean **is sure** it is a big animal.'
Document n°:	WR-P-E-G-0000006651
Corpus example BSD:	*Ik **was er** niet helemaal **zeker van dat** ik niet gevolgd werd.*
	'I **was** not completely **sure that** I was not being followed.'
Document n°:	WR-P-E-G-0000005934
Corpus example CBD:	*Tegen 1995 **was** hij **zeker dat** het virus onschadelijk gemaakt was.*
	'By 1995, he **was sure that** the virus had been defused.'
Document n°:	WR-P-E-G-0000004794

Profile 13:	*(zo)als + su – (zo)als + ob*
Translation:	*like + object*
Corpus example BSD:	*Velen ondergingen hetzelfde lot **als hij**.*
	'Many suffered the same fate **as him**.'
Document n°:	WR-P-E-G-0000006086
Corpus example CBD:	*Vroeger was ik **zoals jou**.*
	'I used to be more **like you**.'
Document n°:	WR-P-E-G-0000000468

Profile 14:	*zodra – van zodra*
Translation:	*as soon as*
Corpus example BSD:	***Zodra** het materiaal er was , legden de metselaars de fundamenten.*
	'**As soon as** the material was there, the masons laid the foundations.'
Document n°:	WR-P-E-G-0000001221
Corpus example CBD:	***Van zodra** ze mij zagen , wisten ze dat ik een natuurtalent was.*
	'**As soon as** they saw me, they knew I was a natural.'
Document n°:	WR-P-E-G-0000000169

Profile 15:	*zulke* + meervoud – *zo'n* + meervoud
Translation:	*such* + plural noun
Corpus example BSD:	*Op **zulke momenten** ben ik één van de gelukkigste mensen die er zijn.*
	'At **such moments**, I am one of the luckiest people in the world.'
Document n°:	WR-P-E-G-0000002324
Corpus example CBD:	***Zo'n mannen** hebben ze nodig.*
	'They need **such men**.'
Document n°:	WR-P-E-G-0000006684

Profile	Belgian Standard Dutch	Colloquial Belgian Dutch	Translation or meaning
1	part + aux + inf (n = 900) aux + inf + part (n = 349)	aux + part + inf (n = 689)	position of the participle in the verbal end group
2	NP + aux +inf (n = 349)	aux + NP + inf (n = 10)	position of the noun phrase in the verbal end group
3	PA + aux + inf (n = 1244)	aux + PA + inf (n = 573)	position of the pronominal adverb in the verbal end group
4	part + inf + inf (n = 113)	inf + part + inf (n = 63)	position of the infinitive in the verbal end group

Table 19. List of the syntagmatic profiles used in case study 2

Profile 1 :	**part + aux + inf – aux + inf + part – aux + part + inf**
Translation:	Position of the participle in the verbal end group
Corpus example BSD:	*Het fort is al enkele malen **ontruimd moeten worden.***
	'The fort **has been evacuated** several times.'
Document n°:	WR-P-E-G-0000006448
Corpus example BSD:	*Sommigen schreeuwden dat Frodo **moest worden afgemaakt.***
	'Some people yelled that Frodo **had to be killed**.'
Document n°:	WR-P-E-G-0000005452
Corpus example CBD:	*De boodschap kwam dat de koolmijnen **zouden afgebouwd worden.***
	'The message came that the coal mines **would be closed down**.'
Document n°:	WR-P-E-G-0000002386

Profile 2:	**NP + aux + inf – aux + NP + inf**
Translation:	position of the noun phrase in the verbal end group
Corpus example BSD:	*We weten alleen dat hij **pintjes kwam drinken** hier in de stationsbuurt.*
	'We only know that he **came** here near the station **to drink beer**.'

Document n°:	WR-P-E-G-0000003826
Corpus example CBD:	*Ik vond het leuk dat ze **kwamen afscheid nemen**.*
	'I liked that they **came to say goodbye**.'
Document n°:	WR-P-E-G-0000007837

Profile 3:	**PA + aux + inf – aux + PA + inf**
Translation:	position of the pronominal adverb in the verbal end group
Corpus example BSD:	*Er is veel waar ze trots **op kunnen zijn***
	'There are a lot of things of **which** they **can be proud**.'
Document n°:	WR-P-E-G-0000006118
Corpus example CBD:	*Maar je weet dat hij er niks **kan aan doen**.*
	'But you know that he **can't do** anything **about** it.'
Document n°:	WR-P-E-G-0000000448

Profile 4:	***part + inf + inf – inf + part + inf***
Translation:	position of the infinitive in the verbal end group
Corpus example BSD:	*Hier onderzoeken wetenschappers welke dierproeven **vervangen kunnen worden** door een alternatief.*
	'Here, scientists investigate which animal experiments **can be replaced** by an alternative.'
Document n°:	WR-P-E-G-0000008298
Corpus example CBD:	*Dat is het eerste wat moet **kunnen gerealiseerd worden**.*
	'That's the first thing that **should be possible to realize**.'
Document n°:	WR-P-E-G-0000003353

5.3   Results and discussion

This section discusses the results of the profile-based correspondence analysis in the second case study. First, we provide some general information about the dispersion of BSD and CBD in the subtitles and we interpret the two-dimensional plot, which visualizes the linguistic choices made by Flemish subtitlers (5.3.1). In the following sections, the influence of the contextual parameters *source language* (5.3.2) and *program genre* (5.3.3) is discussed in more detail in order to generate conclusions with regard to the initial hypotheses.

5.3.1 General observations

Data set	Belgian Standard Dutch		Colloquial Belgian Dutch	
	Absolute n	Relative n (%)	Absolute n	Relative n (%)
Lexical profiles	7052	74	2468	26
Constructional profiles	17061	75	5599	25
Syntagmatic profiles	2955	69	1335	31

Table 20. Overview of the absolute and relative numbers of BSD and CBD attestations per profile set in case study 2

Table 20 visualizes the distribution of BSD and CBD in the subtitles. With regard to the lexical profiles, 74% of the variants are BSD, whereas 26% are CBD lexemes. The constructional profiles show more or less the same proportion: 75% of the variants are BSD constructions, whereas 25% are CBD. For the syntagmatic profiles, the number of CBD variants is somewhat higher (31%), resulting in a lower number of BSD variants (69%). These observations confirm our first hypothesis at least partially, viz. that subtitlers tend to avoid non-standard lexicon. However, this table does not provide us with information about (i) the mutual behavior of the individual language variants and (ii) the contexts in which subtitlers opt for standard or non-standard language. Therefore, we applied profile-based correspondence analysis to visualize the linguistic choices of the subtitlers.
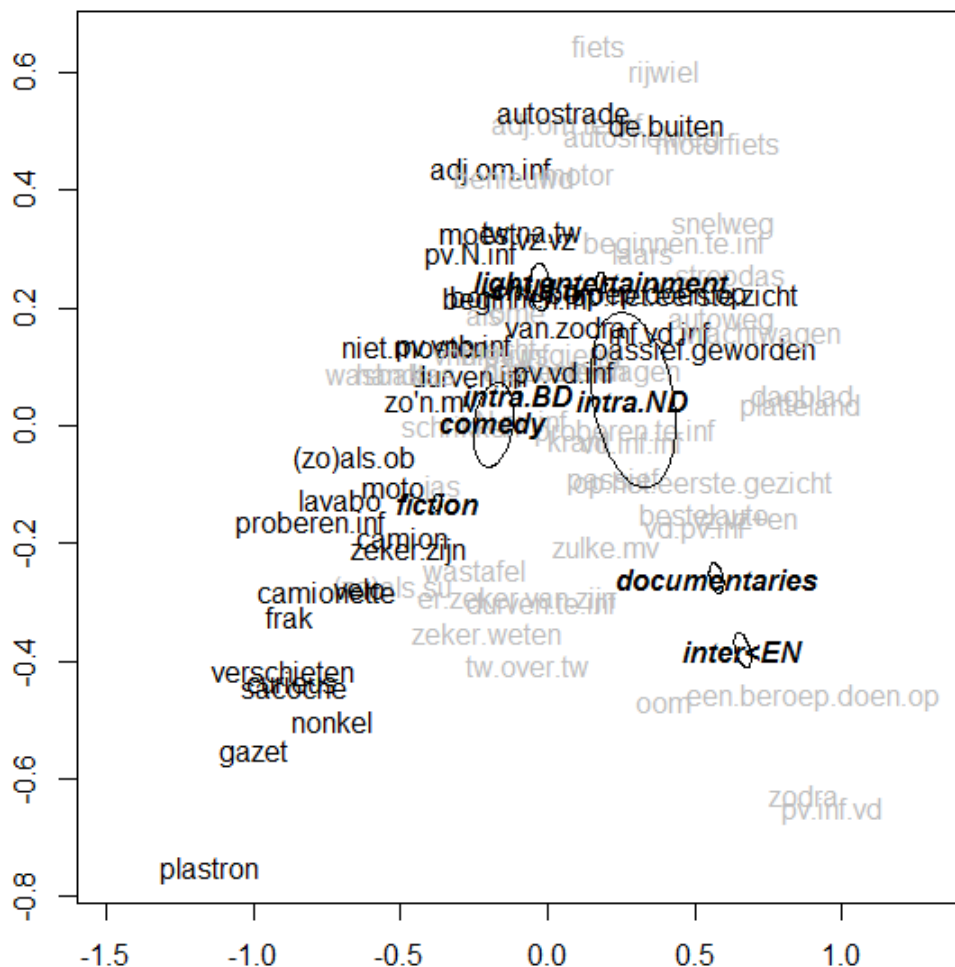
Figure 13. Biplot of the lexical-paradigmatic, constructional-paradigmatic profiles, and syntagmatic profiles, and the contextual parameters *source language* and *program genre* in the subtitle data of the SoNaR Corpus (grey = BSD, black = CBD)

Figure 13 presents the distribution of the lexical, constructional and syntagmatic variants in the subtitles of Belgian-Dutch (*intra.BD*), Netherlandic-Dutch (*intra.ND*), and English (*inter<EN*) speakers on the one hand and five program genres (*child.tv*, *comedy*, *documentaries*, *fiction*, and *light.entertainment*) on the other. The BSD variants are represented in grey and the CBD alternatives are marked in black. If we look at the dispersion of the linguistic variants, we can see that the first dimension (from left to right along the horizontal x-axis) is defined by the dispersion of CBD vs. BSD, since the majority of the black CBD variants are mainly located at the left side, whereas the grey BSD variants are situated at the right side of the plot. Along the second dimension of this plot (from top to bottom along the vertical y-axis), we can observe genre- and source language-related variation, with *children's television* and *light entertainment* at the top of the plot and

interlingual subtitles of English speech at the bottom. The position of the linguistic variants provides information about the relation between subtitlers' choices and the contextual parameters *source language* and *genre*. The closer a contextual parameter is situated toward the CBD variants, the more often CBD is used in the subtitles in that specific language variety or genre. The biplot shows that the linguistic choices of the subtitlers within these contexts differ significantly, since the confidence ellipses do not overlap. In Section 4.3, however, we have already mentioned that correspondence analysis is merely an estimation of the associations in the original data frame. To evaluate the quality of this visualization, we generated a scree plot, which visualizes the accuracy of the plot with respect to the total variation in the data set.
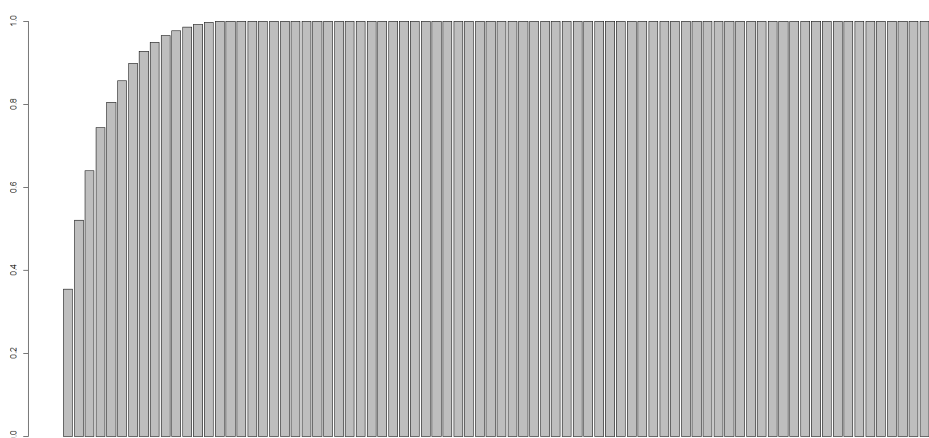
Figure 14. Cumulative scree plot for the distribution of the lexical-paradigmatic, constructional-paradigmatic, and syntagmatic profiles in the subtitle data of the SoNaR Corpus

The scree plot in Figure 14 shows that the degree of representativeness of the two-dimensional plot in Figure 13 is merely 52.13%, with the first dimension (x-axis) capturing 35.53% of the variation and the second axis (y-axis) counting for 16.60% of the variation. If we want to visualize at least 80% of the total variation in the data set, we need a minimum of five dimensions. This presents a practical problem, however, since five-dimensional plots are not easily visualized. In addition, the two-dimensional plot in Figure 13 contains all information of the different profile sets and the contextual parameters together, which yields a chaotic picture. Therefore, we will further focus on the influence of *source language* and *program genre* in the three profile sets separately. By reducing the overload of information, we will not only make the plots more interpretable, the accuracy of the visualization will also increase.

5.3.2　The influence of source language on subtitlers' linguistic choices

Figure 15. (from left to right) Biplot of the lexical-paradigmatic, constructional-paradigmatic profiles, and syntagmatic profiles in the subtitle data of the SoNaR Corpus (grey = BSD, black = CBD)

In Figure 15, which shows the position of the linguistic profiles, it can be observed that the variants are dispersed differently in the three plots. In the plot with the lexical profiles (left plot), most of the grey variants are located at the left side, whereas the black variants are clustered at the right side of the plot. This clearly indicates a divide between the contexts in which BSD variants are used and those in which CBD variants are used, suggesting that subtitlers consciously choose lexical items depending on the context. With regard to the grammatical profiles, a comparable dispersion can be observed for the syntagmatic profiles (right plot), as the grey BSD variants are generally located at the bottom side of the plot and the black CBD variants are clustered at the top of the plot. The plot with the constructional profiles (middle plot), on the other hand, does not clearly divide the grey and the black variants, which implies that in all programs the subtitles contain both BSD and CBD. In other words, the subtitling context (*source language* and *program genre*) hardly has an influence on the use of standard and non-standard constructional variants in the subtitles on Flemish television. This observation partially confirms our second hypothesis, viz. that the use of standard and colloquial grammatical constructions is less context dependent than the use of standard and colloquial lexemes.
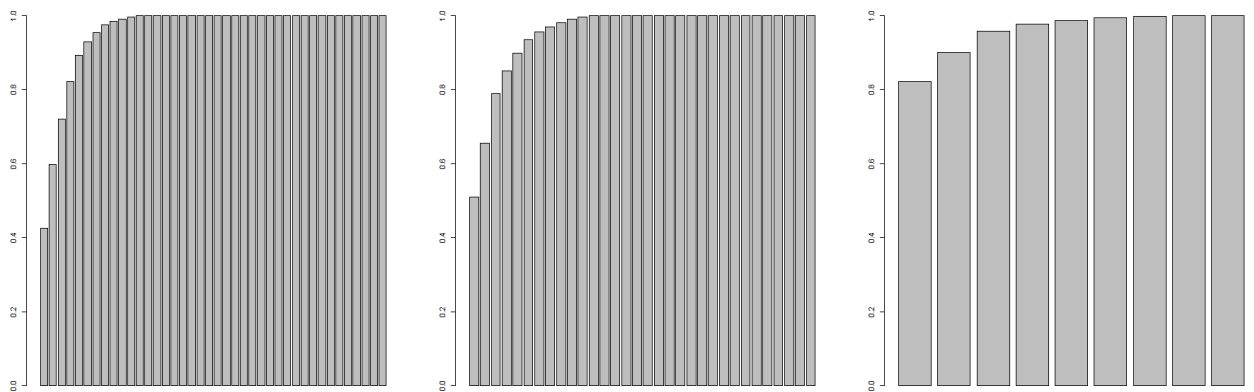
Figure 16. (from left to right) Cumulative scree plots for the distribution of the lexical-paradigmatic, constructional-paradigmatic, and syntagmatic profiles in the subtitle data of the SoNaR Corpus

If we take a look at the scree plots for the lexical, the constructional and the syntagmatic profiles in Figure 16, we see that the two-dimensional visualizations represent respectively 59.73%, 65.49%, and 89.94% of the language variation in the three data sets. In other words, the accuracy of the visualization has increased compared to the scree plot in Figure 14, which counted for merely 52.13% of the total variation. Still, four dimensions are required to reach 80% in the lexical and the constructional plot. This presents us with a practical problem, however, as four-dimensional plots are difficult to visualize. Although we are aware that the visualization in Figure 15 counts for merely 59.73 % of

the variation in the lexical data and 65.49% of the variation in the constructional data, we will use this plot for the exploratory analysis of the language variation in the subtitles.
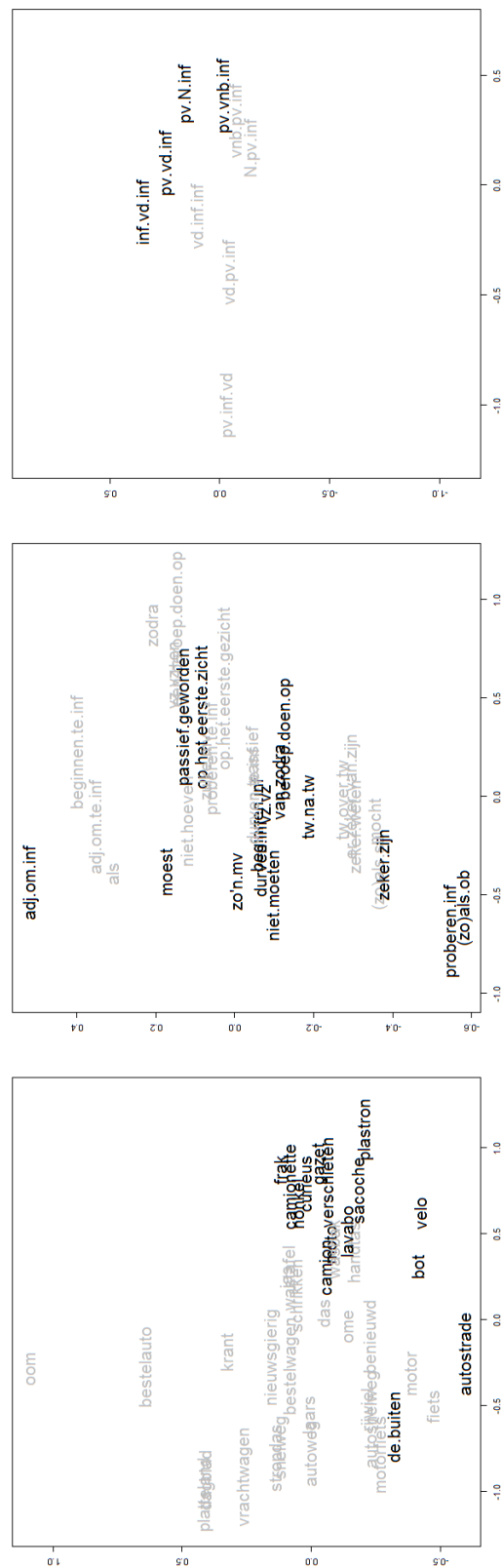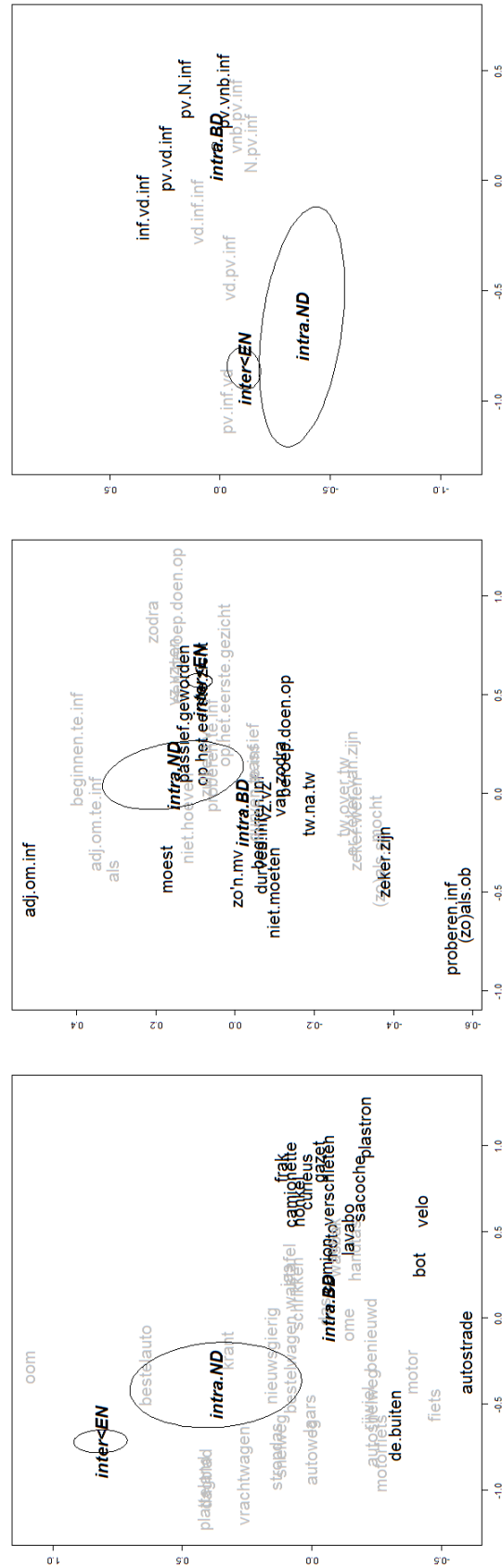
Figure 17. (from left to right) Biplot of the lexical-paradigmatic, constructional-paradigmatic, and syntagmatic profiles, and the source language varieties in the subtitle data of the SoNaR Corpus (grey = BSD, black = CBD)

Since we want to verify whether the original speech in the television program determines the linguistic choices of the subtitlers, the different source language varieties are plotted upon the lexical, the constructional and the syntagmatic profiles in Figure 17. The source language is either Belgian Dutch (*intra.BD*), English (*inter<EN*) or Netherlandic Dutch (intra.ND). The resulting biplots reveal some interesting findings. First, when we look at the position of the different source language varieties relative to the dispersion of the linguistic variants, it can be observed that the linguistic choices made in each of these contexts are significantly different ($p < .05$) in the lexical and the constructional data set, since the ellipses do not overlap. In the syntagmatic set, however, there is an overlap between the ellipses of *intra.ND* and *inter<EN*, which means that the subtitlers' syntagmatic choices are not significantly different when subtitling a Netherlandic-Dutch or an English speaker. Second, it is clear that the interlingual subtitles of English television programs are located close to the BSD variants in the three plots, which implies that in this context subtitlers use standard language to a large extent. However, this association with the BSD variants is less outspoken for the constructional profiles, since the distance from the *inter<EN* subtitles to the black CBD variants is smaller there. Similarly, the subtitles of Netherlandic-Dutch programs are clearly related to the BSD variants for the lexical and syntagmatic profiles, whereas for the constructional profiles this source language variety is also surrounded by some CBD variants. Finally, the intralingual subtitles of Belgian-Dutch programs contain a lot of non-standard language, since this source language variety is located close to the CBD variants in the three plots. These findings verify hypothesis 3d and simultaneously confirm the results of our previous case study, viz. that subtitlers use more CBD when subtitling Belgian-Dutch speech compared to Netherlandic-Dutch and English speech. The lower level of norm adherence in intralingual subtitles of Belgian-Dutch speakers can be explained by taking into account that in this context, subtitlers are directly exposed to original Belgian-Dutch speech. As already mentioned, CBD variants occur very frequently in spoken language on Flemish television (cf. Saman 2003; Van Gijsel 2008; Zenner et al. 2009). Consequently, it is to be expected that the subtitles of these programs also contain a high amount of CBD, because subtitlers plausibly transfer the CBD variants in the original Belgian-Dutch speech to the subtitles. In other words, the higher degree of non-standard language in intralingual subtitles of Flemish speakers is most likely caused by direct interference from the language use in the original Belgian-Dutch television program. However, we were not able to consult the original footage of the television programs, since the SoNaR Corpus does not contain the spoken source texts. In case study 3, we will try to substantiate these assumptions by using a specialized parallel corpus.

5.3.3 The influence of (program) genre on subtitlers' linguistic choices

Further analyses revealed that not only the contextual parameter *source language* but also *program genre* influences the subtitlers' linguistic choices. Especially the intralingual subtitles of Belgian-Dutch television programs (and, to a lesser extent, the intralingual subtitles of Netherlandic-Dutch television programs) show significant genre variation. The program genre does not influence the linguistic choices made in the interlingual subtitles of English programs, since these subtitles mainly contain standard language (cf. section 5.3.2.). In this section, we discuss the influence of the program genre on the linguistic choices of the subtitlers in more detail.

Figure 18. (from left to right) Biplot of the lexical-paradigmatic, constructional-paradigmatic profiles, and syntagmatic profiles, and the *program genres* varieties in the subtitle data of the SoNaR Corpus (grey = BSD, black = CBD)

The linguistic choices made in each of the genres are significantly different (p < .05) in the lexical and the constructional data sets, since the ellipses do not overlap. In the syntagmatic plot, however, there is only a significant difference between the subtitles of *documentaries* and the other genres. The ellipses of *light entertainment*, *children's television*, *fiction*, and *comedy* overlap, which implies that the subtitlers' linguistic choices in these genres are not significantly different. Further, the position of the different genres vis-à-vis the linguistic variants shows that there is a lot of genre variation, which is particularly caused by the influence of the Belgian-Dutch spoken programs in our data set. The English and Netherlandic-Dutch spoken programs show (almost) no variation (cf. supra). First, in the plot with the lexical profiles (left plot), it is clear that *fiction* and *comedy* are most strongly related to CBD, as these genres are situated in the core of the colloquial variants. In contrast, subtitles in *light entertainment* are more closely related to the BSD variants, although this genre is still surrounded by some of the CBD variants, whereas the subtitles in *documentaries* and *children's television* are most strongly related to the BSD variants, their distance to the non-standard variants being larger. Second, in the plot with the constructional profiles (middle plot) the program genre has hardly any influence on the use of standard and colloquial constructions in the subtitles on Flemish television. This observation confirms once again that the use of standard and non-standard grammatical constructions is less context-dependent than the use of standard and non-standard lexemes (verification of hypothesis 2). Finally, in the plot with the syntagmatic profiles (right plot) the genre *light entertainment* is located closest to the CBD variants, whereas *comedy* and *fiction* are located closer to the BSD variants. *Documentaries* and *children's television*, on the contrary, are the most norm-adhering genres, since the distance from these genres to CBD variants is the largest. Furthermore, if we compare the three data sets, we notice that, on the one hand, the relative distance of *children's television* to the CBD variants is smaller for the constructional and syntagmatic profiles than for the lexical profiles. On the other hand, the relative distance of *light entertainment* to the core of the CBD variants is larger for the lexical profiles than for the constructional and syntagmatic profiles.

The main conclusions that can be drawn from these findings are the following. First, the results demonstrated that the contextual parameter *program genre* affects the linguistic choices of subtitlers. The analyses revealed that subtitles in *documentaries* and *children's television* mainly contain standard language, whereas subtitles in *fiction* and *comedy* contain a lot of non-standard language and subtitles in *light entertainment* take up a middle position (verification hypothesis 3). Simultaneously, it was shown that subtitlers more often avoid non-standard lexemes than non-standard constructional and syntagmatic variants in *documentaries* and *children's television* (cf. appendix 3 for the total number of BSD and CBD attestations per data set in each genre). This conclusion confirms our first hypothesis, viz. that Flemish subtitlers are more norm-adhering toward lexical variants than toward constructional and syntagmatic variants. A possible explanation for

this outcome is that lexical features are more salient than grammatical constructions (cf. Lybaert 2014b), as a result of which subtitlers are more aware of the standard or non-standard character of the former than of the latter.

So far, the analyses have revealed in which genres subtitlers tend to use BSD and in which genres they tend to avoid it. However, we also want to investigate which specific features of these genres (*program purpose*, *target audience* and *cast*) determine the subtitlers' linguistic choices. What we still do not know, for instance, is (i) whether subtitles contain more non-standard constructional variants than non-standard lexical and syntagmatic variants in entertainment programs and (ii) whether actors are subtitled differently compared to non-actors. To answer these questions, we calculated and visualized the relative distances between the subparameters and the linguistic profiles.

5.3.3.1 The influence of program purpose on subtitlers' linguistic choices

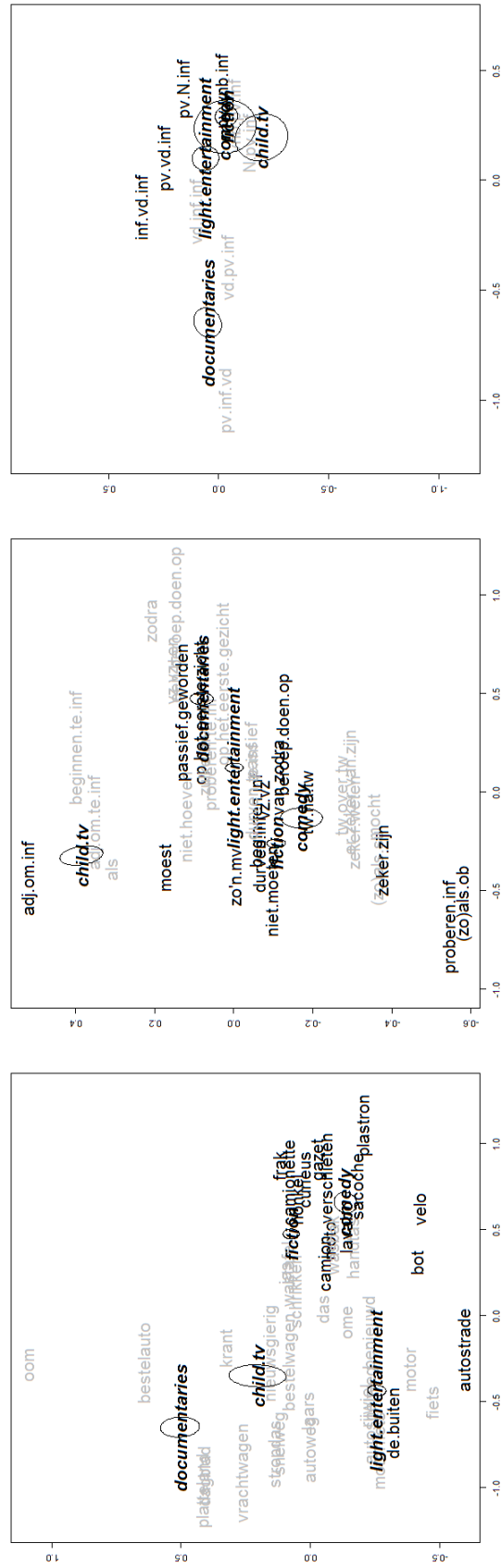

Figure 19. (from left to right) Biplot of the lexical-paradigmatic, constructional-paradigmatic, and syntagmatic profiles, and the *program purpose* in the subtitle data of the SoNaR Corpus (grey = BSD, black = CBD)

Figure 19 is basically the same as Figure 18, but here it represents the position of the different *program purposes* relative to the position of the linguistic variants. With regard to the lexical profiles (left plot), the subtitlers' linguistic choices are significantly different in the four subtitling contexts. It can immediately be observed that subtitles of *laughing* programs are most clearly related to CBD variants, whereas subtitles of *informing* programs are most related to BSD variants. Subtitles of *infotaining* and *entertaining* programs are also related to the BSD variants, although less notably so than *informing* programs, as they are still surrounded by some CBD variants. As mentioned above, the use of standard vs. non-standard language is less context-dependent for the constructional profiles (middle plot) than for the lexical (left plot) and syntagmatic (right plot) profiles, since the position of the constructional CBD and BSD variants is extremely varied. Based on the position of the ellipses, it is clear that the linguistic choices of the subtitlers in *informing* and *infotaining* programs, on the one hand, and *laughing* and *entertaining* programs, on the other, are similar for the constructional profiles. In the plot with the syntagmatic variants, the program purpose affects the linguistic choices of the subtitlers less, since the variation between the different contexts is less prominent. It can be observed that subtitles in *informing* programs contain more BSD variants than subtitles in *infotaining*, *entertaining*, and *laughing* programs, which are located much closer to the CBD variants.

In sum, the analyses showed again that in some television programs subtitlers more often avoid the use of colloquial lexemes than colloquial constructions (verification of hypothesis 1). Furthermore, the results demonstrated that the program purpose influences the linguistic choices of subtitlers. The lower degree of norm adherence in non-informative television programs (*entertaining* and *laughing*) can be explained by their general aim. The main objective of entertaining and humorous programs is to amuse the audience, thus creating an informal, spontaneous atmosphere, which has a greater chance of showing spontaneous, colloquial utterances (cf. McIlvenny et al. 1992; Remael 2003; Rutter 1997). As a consequence, it seems plausible that subtitlers (consciously or unconsciously) reproduce this spontaneous nature of the television program by using non-standard, colloquial features in the subtitles (verification of hypothesis 3a). We already mentioned that the SoNaR Corpus does not contain the original audio fragments, so we were not able to consult the original spoken source text. Therefore, we will set up a third case study which compares the original speech in the TV programs to the corresponding intralingual subtitles (cf. Chapter 6)

5.3.3.2 The influence of target audience on subtitlers' linguistic choices
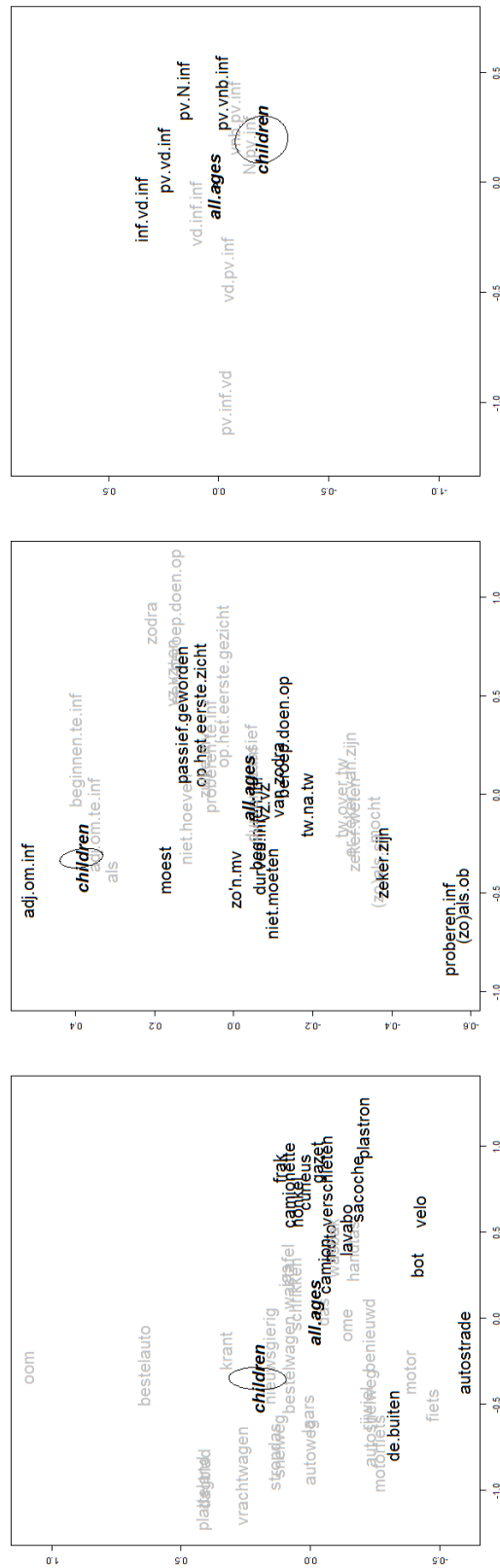
Figure 20. (from left to right) Biplot of the lexical-paradigmatic, constructional-paradigmatic, and syntagmatic profiles, and the *target audience* in the subtitle data of the SoNaR Corpus (grey = BSD, black = CBD)

In Figure 20, the influence of the target audience is visualized. The most interesting observation is that subtitles in television programs intended for *children* are clearly related to the BSD variants in the lexical (left plot) and the syntagmatic plot (right plot), whereas these subtitles also contain colloquial variants in the constructional plot (middle plot). Subtitles in television programs without a target age group (*all.ages*) are generally more related to the CBD variants, although their closer position is less notable for the syntagmatic profiles. In the lexical plot, however, the relative distance of *children* to the CBD lexemes is larger than the relative distance of *all.ages* to the CBD lexemes, which implies that the subtitles of children's television programs particularly contain standard language at lexical level.

The main conclusion emerging from Figure 20 is that the influence of the target audience on the subtitlers' linguistic choices is generally not outspoken, except for the lexical profiles. The analyses demonstrated that subtitlers more often tend to avoid non-standard lexemes in television programs intended for *children* (cf. appendix 3 for the total number of BSD and CBD attestations per data set for target audience). Furthermore, subtitles in children's television programs generally contain more standard language than subtitles in programs for all ages (verification of hypothesis 3b). This could be explained by the educational footing of children's programs which aim to perform an exemplary role, also on the level of language use. As VRT explicitly prescribes the use of standard language in children's television programs (Hendrickx 1998), it is plausible that the subtitlers reproduce these BSD lexemes. For similar reasons, it can be assumed that subtitlers tend to convert the spoken colloquialisms into standard language. When comparing the original speech of the TV program to the corresponding intralingual subtitles in case study 3, we will try to substantiate these assumptions.

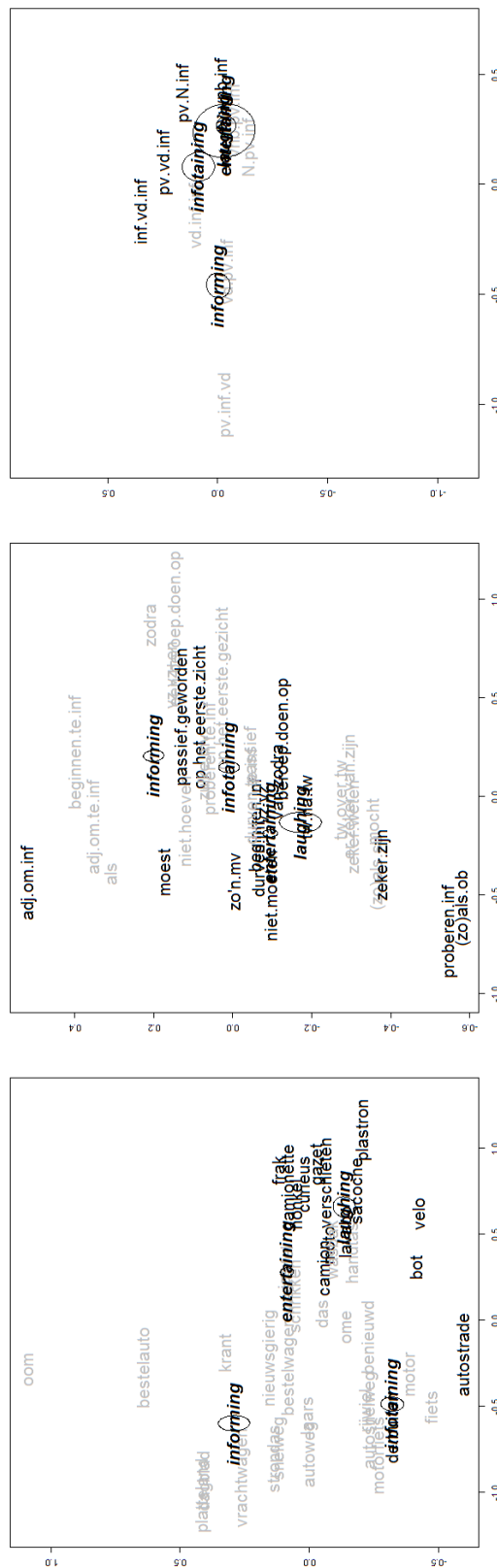## 5.3.3.3   The influence of cast on subtitlers' linguistic choices
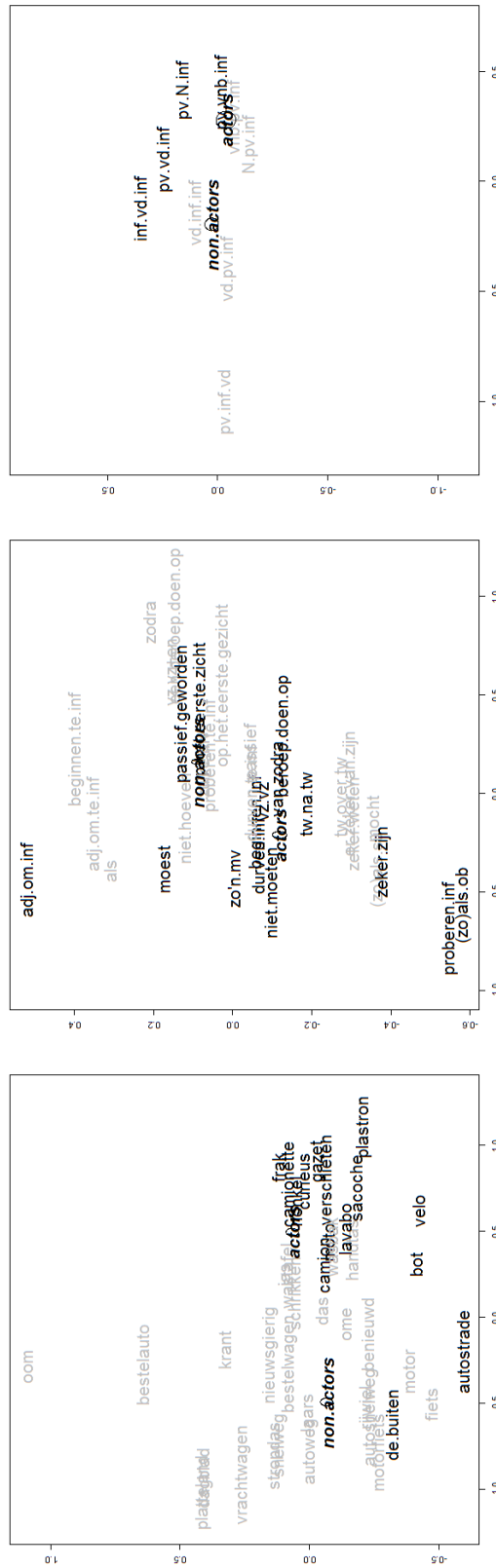
Figure 21. (from left to right) Biplot of the lexical-paradigmatic, constructional-paradigmatic, and syntagmatic profiles, and the *cast* in the subtitle data of the SoNaR Corpus (grey = BSD, black = CBD)

Figure 21 visualizes the influence of cast on the lexical (left plot), the constructional (middle plot) and the syntagmatic (right plot) profiles. The three plots show that the linguistic choices made in subtitles of *actors* and *non-actors* are significantly different, as the ellipses do not overlap. Subtitles of *actors* are most related to the CBD variants in both the lexical and syntagmatic plot, whereas subtitles of *non-actors* are located closer to the BSD variants. Especially in the lexical plot, the relative distance of *non-actors* to the CBD lexemes is larger than the relative distance of *actors* to the CBD lexemes, which implies that subtitles of *non-actors*, in particular, contain standard language at the lexical level. For the constructional profiles, however, subtitles of *actors* and *non-actors* both contain standard and non-standard language, since they are surrounded by BSD variants as well as by CBD variants.

In other words, the analyses revealed that cast is an important factor in the linguistic choices of subtitlers. Contrary to what was assumed in hypothesis 3c, subtitles of *non-actors* tend to contain standard language, especially for the lexicon, whereas subtitles of *actors* contain a lot of non-standard language. A similar 'linguistic hierarchy' was observed in Remael et al. (2008). In their study, it was demonstrated that none of the television hosts were subtitled, whereas all of the interviewees were. This implies that, unlike television hosts, the language use of interviewees is expected to contain a lot of colloquial language and be difficult to be understand by the viewers and, as a consequence, needs to be 'translated into Standard Dutch' in the subtitles. This could explain why the subtitles of *non-actors* (e.g. interviewees) contain more BSD than those of *actors*. Furthermore, Remael (2003: 226) emphasized that 'it is important to distinguish the scripted dialogue (of *actors*) [my addition] of fiction films or TV series from the more or less spontaneous speech of a live interview (*non-actors*)' [my addition], because 'in fictional dialogue both register and the interactional features of conversation are part of a carefully constructed narrative that also relies on other sign systems to communicate with the viewer'. In other words, the use of non-standard language in the subtitles of *actors* could be a conscious strategy in the communication system of the subtitlers. After all, the public broadcaster is more tolerant toward the use of non-standard language to maintain the authenticity of the characters in entertainment programs (Hendrick 1998), so it can be expected that the colloquial features in the original speech of the *actors* are transferred to the subtitles. In that way, the subtitlers want to avoid a situation where 'the characters speak like a printed page' (Rosa 2001: 216). Furthermore, it was again demonstrated that subtitlers more often avoid non-standard lexemes in subtitles of *non-actors* (cf. appendix 3 for the total number of BSD and CBD attestations per data set for cast).

5.4   Concluding remarks

Building on a corpus of subtitles that were produced by the Flemish public broadcaster between 2000 and 2005, we statistically analyzed whether subtitlers use more lexical colloquialisms than grammatical colloquialisms. Furthermore, it was investigated which contextual factors (*source language* and *program genre*) affect these lexical and grammatical choices. The results demonstrated that, even though VRT's subtitling guidelines 'more or less' allow the reproduction of colloquial lexicon (Dewulf & Saerens 2000), Flemish subtitlers are more norm-adhering for lexemes than for grammatical constructions. More specifically, non-standard lexemes occur less frequently in certain subtitling contexts on Flemish television than non-standard grammatical constructions. The most obvious explanation offered for these results is that CBD lexical features are significantly more often perceived by the subtitlers than CBD syntactic features are (Lybaert 2014b). Subtitlers consequently tend to convert these salient lexical colloquialisms into BSD, since they want to avoid the risk of ignoring the subtitling guidelines. On the contrary, non-standard constructions probably pass by unnoticed and are unconsciously reproduced in the subtitles. Further, in-depth research into salience differences between language variants will be performed in the next case study in order to provide a more fine-grained insight in the way in which salience affects subtitlers' linguistic choices.

Second, it was shown that the source language and the program genre are two factors that cause norm-related differences in the subtitle corpus. First, subtitlers are more norm-adhering when subtitling English or Netherlandic-Dutch spoken television programs compared to Belgian-Dutch spoken programs. In other words, the subtitles contained significantly more CBD variants when the source language of the original footage was (Colloquial) Belgian Dutch. Second, the language in the intralingual subtitles in Belgian-Dutch spoken programs was to a large extent influenced by the program genre. The subtitles in more informative programs (documentaries) and programs intended for children (children's television) mainly contain standard language, whereas the number of CBD lexemes and constructions increased significantly in the subtitles of humorous (comedy) and entertainment (fiction) programs. Light entertainment took a middle position.

A plausible explanation for these results is that subtitlers directly transfer the BSD and CBD variants of the original footage to the subtitles. On the one hand, VRT adopts an exemplary role with regard to the language used in informative programs (e.g. documentaries) and television programs intended for children (Hendrickx 1998). In these genres, the Language Charter explicitly prescribes the use of standard language. As a result, it is plausible that the subtitlers reproduce the BSD variants that are used in these television programs. On the other hand, 'the relationship between standard and non-standard language plays a central role in the creation of dialogue and in the credibility of

the participants' stories' (Hedin 2009: 42). In VRT's Language Charter, the use of tussentaal is therefore occasionally allowed: 'Dialect and tussentaal can be used in Flemish soaps, serials and comedy series to improve the credibility of the characters' [my translation] (Hendrickx 1998: 2). Furthermore, the use of non-standard language in television dialogue can be a conscious strategy to entertain the audience or to create a comic effect (e.g. McIlvenny et al. 1992; Remael 2003; Rutter 1997). In this context, several studies have shown that the use of tussentaal in Flemish entertainment and comedy programs can be functional (e.g. De Ridder 2007; Lefevere 2011; Saman 2003; Van Hoof 2015). As a consequence, we can assume that the subtitles in these program genres contain a lot of CBD because these colloquial features are also used in the original spoken source text. It should not surprise us that subtitlers want to maintain the 'linguistic effect' of the original footage and that they, thus, reproduce the spoken colloquialisms in the subtitles. However, since we were not able to analyze the original, spoken television fragments, the next case study will compare the original Belgian-Dutch speech in twenty television programs to the corresponding intralingual subtitles in order to verify the aforementioned assumptions. Nevertheless, the results of the present case study emphasize the importance of contextual factors that should be taken into account when analysing subtitlers' linguistic choices, since it was pointed out that these choices are largely influenced by the purpose of the television program and the target audience.

Chapter 6
Case study 3: analyzing the influence of the Belgian-Dutch spoken source text on the corresponding intralingual subtitles

This third case addresses the question to what extent Flemish subtitlers reproduce the Belgian-Dutch colloquialisms from the spoken source text in the subtitles. By comparing the language used in twenty television programs to the corresponding intralingual subtitles, it will be examined whether the subtitlers more often opt for lexical colloquialisms than morphological or syntactic colloquialisms, and whether the program genre influences these linguistic choices. First, the ratio between Colloquial Belgian Dutch (CBD) and Belgian Standard Dutch (BSD) in the subtitle corpus in relation to the use of CBD in the spoken corpus is calculated. In addition, subtitlers were interviewed to get more contextual information about the practical as well as the political concerns they have to deal with and to find out why they opt for the reproduction or translation of certain Belgian-Dutch colloquialisms.

In the present case study, we will consult a more recent corpus (2014-2016) than in case studies 1 and 2, which involves that these subtitles were produced by the public broadcaster almost 15 years later than the subtitles in the SoNaR corpus. In the meantime, VRT has updated its subtitle guidelines (VRT 2009), which show more tolerance (and even promote) the use of colloquial lexicon in television subtitles. With regard to colloquial grammatical constructions, VRT's new guidelines do not allow the reproduction of morphosyntactic colloquialisms. The style guide even includes a number of colloquial constructions that must be converted into standard language, which increases the sensitivity toward colloquial grammar.

6.1 Hypotheses

Considering VRT's subtitling guidelines and building on the results of the first and the second case study, we formulate following hypotheses.

➢ **Hypothesis 1:** the CBD features from the spoken source text are more often translated into BSD than reproduced in the subtitles. In line with the increased use of tussentaal on Flemish television, VRT stimulates its subtitlers to reproduce the CBD features in order to retain the authenticity of the television program and the characters. However, the main function of subtitling is to support the audience by improving the intelligibility of the program. As the abundant use of colloquial variants could affect the intelligibility of the subtitles, it can be expected that Flemish subtitlers tend to restrict the reproduction of these colloquial features to a minimum.

➢ **Hypothesis 2:** subtitles on Flemish television contain more CBD lexemes than CBD morphological or syntactic constructions. Although the results of the previous chapter have shown that subtitlers are more norm-adhering toward CBD lexemes than toward CBD grammatical constructions, we expect an opposite outcome in this case study. This can be attributed to the recent corpus (2014-2016) that will be consulted and the revised version of VRT's style guide (VRT 2009). These guidelines prescribe that colloquial lexicon must be reproduced as much as possible. On the contrary, colloquial morphosyntactic constructions should be converted into standard language. Given VRT's tolerance toward the use of colloquial lexemes and its preference for correcting grammatical colloquialisms, it can be expected that subtitlers use CBD lexicon to a large extent, whereas they tend to avoid the use of colloquial grammatical variants.

➢ **Hypothesis 3:** the number of CBD variants in the subtitles will be higher in entertainment, infotainment and humor programs (vs. informative programs). As the public broadcaster allows the use of non-standard varieties in these genres (Hendrickx 1998), it can be assumed that more colloquial language will occur in non-informative programs than in informative programs. As a consequence, subtitlers will be more likely to reuse the spoken colloquialisms in entertainment, infotainment and humor programs, whereas they will tend to eliminate these colloquialisms in informative genres (cf. case study 2). In addition, we expect that these assumptions, together with the results of the quantitative analyses, will be confirmed by the qualitative part of this case study.

6.2 Variable selection

In order to investigate to what extent subtitlers' linguistic choices are influenced by the original footage of the television program, we calculated how often Flemish subtitlers reproduce the spoken colloquialisms from the original source text rather than converting these colloquial variants into standard language. Therefore, we selected 55 CBD items which were divided into three types: lexical, morphological, and syntactic features. These colloquial variants were extracted from our spoken corpus, together with the corresponding alternatives in the subtitle corpus to verify whether the subtitlers opt for the reproduction of the colloquialisms or whether they convert the colloquialisms into standard language. In addition, we examined whether the subtitlers also convert BSD features from the spoken source text into CBD variants in the subtitles. Again, normative sources were consulted to verify the status of these linguistic features. For the lexical items, we only selected features that were labelled either as colloquial or as standard language in the latest edition of Van Dale dictionary (Van Dale et al. 2015). The morphological and syntactic features are based on *Taaladvies* and a list of *tussentaal* features of De Caluwe (2006) and language advisor Hendrickx (2001). We only included a feature in the list if these sources agreed in characterizing it as standard or non-standard language. As we already mentioned in Section 3.2.3, the status of a couple of these features is nowadays disputed. The possessive construction with *zijn*, for instance, is accepted by some sources in informal written language. Nevertheless, we have added these items to our list of colloquial constructions, as they are still not considered BSD.

Furthermore, we added to our data set nine colloquial features that were not used in the subtitles, but which are also considered typical ingredients of CBD. Although these features did not occur in the subtitles, it could be interesting to know how often subtitlers convert these typical CBD elements into BSD. Therefore, we consulted lists of tussentaal features, compiled by several Dutch linguists in the past few decades (De Caluwe 2006; Everaert 1998; Geeraerts et al. 2000; Hendrickx 2001; Lebbe 1996; Taeldeman 2008; Van Gijsel 2008).

The data extraction resulted in a final data set of 1,756 relevant CBD attestations (spoken: *n* = 1,616; subtitles: *n* = 140) and 1,476 BSD alternatives in the subtitles. Tables 21 to 24 provide an overview of the number of attestations of each CBD feature in both the spoken and the subtitle corpus. In addition, a detailed description of the features is given underneath each table. The example sentences for Table 21, 22, and 23 were extracted from the subtitle corpus, whereas the example sentences for the additional set of colloquial features were extracted from the spoken corpus, since these items did not occur in the subtitles. Furthermore, each example is given a code that refers to the original corpus document from which the sentence was extracted.

Feature	Colloquial Belgian Dutch	Translation	Attestations	
			Spoken	Subtitles
1	accident	traffic accident	1	1
2	ajuin	onion	2	2
3	ambras	quarrel	1	1
4	appelsien	orange	4	3
5	boeleke	pet name for a baby	2	2
6	afbollen	get out	1	1
7	brol	trash	3	3
8	buizen	to flunk	6	6
9	camionette	delivery van	1	1
10	chance	luck	2	1
11	chapelure	breadcrumbs	1	1
12	chichi madam	chichi lady	1	1
13	dagdagelijks	daily	1	1
14	efkes	just (temporal)	41	1
15	flik	cop	3	3
16	fretten	to scoff (food)	4	4
17	frigo	fridge	1	1
18	in het hol van Pluto	at the back of beyond	1	1
19	gelijk	like (comparison)	10	1
20	kostelijke affaire	expensive deal	1	1
21	kozijn	cousin	1	1
22	kuisen	to clean	5	5
23	kuisvrouw	cleaning lady	1	1
24	madam	madam	4	3
25	nonkel	uncle	4	4
26	omwille van	because of	3	3
27	patat	patato	4	4
28	plezant	cheerful	6	6
29	saucisse	sausage	1	1
30	schoon	good-looking	13	7
31	seffens	later	6	1
32	sjotten	play soccer	1	1
33	smossen	to make a mess of	1	1
34	stoefen	to brag	1	1
35	vijzen	to screw	1	1
36	weeral	again	1	1
37	eens	as soon as	2	2

38	*zot*	*crazy*	5	5
39	*zever*	*twaddle*	3	3
40	*zwanzen*	*to joke*	2	2

Table 21. List of the lexical features that were used in case study 3

Feature 1:	*accident*
Translation:	*traffic accident*
Corpus example:	*Allee, ik heb precies een **accident** gehad.*
	'Gee, it is like I had an **accident**.'
Document n°:	SULE1.2

Feature 2:	*ajuin*
Translation:	*onion*
Corpus example:	*Onze papa deed dat ook met **ajuin**.*
	'Our dad did that with **onion** too.'
Document n°:	SULE3.2

Feature 3:	*ambras*
Translation:	*quarrel*
Corpus example:	*Hoe? Is ze **ambras** komen maken?*
	'What? Did she come here to pick a **quarrel**?'
Document n°:	SUF3.2

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.'
Document n°:	SULE3.1

Feature 5:	*boeleke*
Translation:	*pet name for a baby*
Corpus example:	*Waar is dat klein **boeleke** hier?*
	'Where is the little **baby**?'
Document n°:	SUC2.1

Feature 6:	*afbollen*
Translation:	*get out*
Corpus example:	***Bol het af**, jong.*
	'**Get out**, you.'
Document n°:	SUC3.1

Feature 7: *brol*
Translation: *trash*
Corpus example: *600 euro aan **brol** die we niet nodig hebben.*
'600 euros of **trash** that we don't need.'

Document n°: SUC4.2

Feature 8: *buizen*
Translation: *to flunk*
Corpus example: *Maar je bent **gebuisd** voor lo, Kleine.*
'But you **flunked** PE, little one.'

Document n°: SUC4.2

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**.'

Document n°: SUF4.2

Feature 10: *chance*
Translation *luck*
Corpus example: ***Chance** dat dat niet gevallen is.*
'**Luckily** it didn't fall'.

Document n°: SULE3.1

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.'

Document n°: SULE3.2

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.'

Document n°: SUF3.1

Feature 13: *dagdagelijks*
Translation: *daily*
Corpus example: *..die wij op school vaak op **dagdagelijkse** basis meemaken.*
'...which happens **daily** at school.'

Document n°: SUC1.1

Feature 14:   *efkes*

Translation: *just (temporal)*

                       *Heel **efkes**....*

Corpus example: '**Just** a minute...'

Document n°:   SULE3.2

Feature 15:   *flik*

Translation: *cop*

Corpus example: *Daar, de **flikken**.*

                       'There are the **cops**.'

Document n°:   SUC4.1

Feature 16:   *fretten*

Translation: *to scoff (food)*

Corpus example: *Ik **fret** chips.*

                       'I **scoff** chips.'

Document n°:   SUC2.2

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.'

Document n°:   SUD4.2

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**.'

Document n°:   SUF1.2

Feature 19:   *gelijk*

Translation: *like (comparison)*

Corpus example: ***Gelijk** Paulien en Ruben.*

                       '**Like** Paulien and Ruben.

Document n°:   SUF3.2

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.'

Document n°:   SUC2.1

Feature 21: *kozijn*

Translation: *cousin*

*Onze pa laat zijn **kozijn** daarnaar kijken.*

Corpus example: 'Our dad will show it to his **cousin**.'

Document n°: SUF4.2

Feature 22: *kuisen*

Translation: *to clean*

Corpus example: *Om het huis te **kuisen** en te koken.*

'To **clean** the house and to cook.'

Document n°: SULE1.2

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.'

Document n°: SULE1.2

Feature 24: *madam*

Translation: *madam*

Corpus example: *Maar **madam**, jij moet je toch niet excuseren.*

'But **madam**, you don't have to apologize.'

Document n°: SUF4.1

Feature 25: *nonkel*

Translation *uncle*

Corpus example: *Dat is het nummer van mijn **nonkel**.*

'That is the number of my **uncle**.

Document n°: SUF4.2

Feature 26: *omwille van*

Translation: *because of*

Corpus example: *..niet zozeer **omwille van** een gebrek aan startkapitaal..*

'..not so much **because of** a lack of starting capital..'

Document n°: SUD2.2

Feature 27: *patat*

Translation: *patato*

Corpus example: *Een **patat** zonder zout...*

'A patato without salt...'

Document n°: SULE3.2

Feature 28:   *plezant*

Translation: *cheerful*

*Wobbe, pépé, ik vond het super**plezant**.*

Corpus example: 'Wobbe, gramps, I had so much fun.'

Document n°:   SUCT3.2

Feature 29:   *saucisse*

Translation: *sausage*

Corpus example: *En dat is met stukjes **saucisse** erin, ik denk Boulogne.*

'There are some pieces of **sausage** in it, I think it is Boulogne.'

Document n°:   SULE3.1

Feature 30:   *schoon*

Translation: *good-looking*

Corpus example: *Zo **schoon** was die toch niet.*

'He was not that **good-looking**.'

Document n°:   SUC2.1

Feature 31:   *seffens*

Translation: *later*

Corpus example: *En dan kunnen we **seffens**, als ze in de oven steken..*

'And **later**, when they are in the oven, we can .....'

Document n°:   SULE3.2

Feature 32:   *sjotten*

Translation: *play soccer*

Corpus example: *Mijn vlees is aan het rusten. **Sjotten**?*

'The meat is resting. Let's **play soccer**'

Document n°:   SULE3.1

Feature 33:   *smossen*

Translation: *to make a mess of*

Corpus example: *Ben je aan het **smossen**?*

'Are you **making a mess of** it?'

Document n°:   SULE1.2

Feature 34:   *stoefen*

Translation: *to brag*

Corpus example: *Ik ben dat **gestoef** van Toon beu.*

'I am tired of Toon's **bragging**.'

Document n°:   SUF3.1

Feature 35: *vijzen*

Translation: *to screw*

*Kom, **vijs** die deur aan de kast en hou je bakkes.*

Corpus example: 'Come on, **screw** this door onto the closet and shut up.'

Document n°: SUF3.2

Feature 36: *weeral*

Translation: *again*

Corpus example: *Wat? **Weeral**?*

'What? **Again**?'

Document n°: SUF3.1

Feature 37: *eens*

Translation: *as soon as*

Corpus example: ***Eens** die basisbehoeften vervuld zijn..*

'**As soon as** these basic needs are fulfilled..'

Document n°: SUD4.1

Feature 38: *zot*

Translation: *crazy*

Corpus example: Zot, ik ben zes keer gebuisd hè.

'Are you crazy? I flunked six times.'

Document n°: SUC4.2

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**.'

Document n°: SUF3.2

Feature 40: *zwanzen*

Translation: *to joke*

Corpus example: *Ik **zwans** maar, hè.*

'Oh, I am **joking**.'

Document n°: SUC1.1

Feature	Colloquial Belgian Dutch	Translation	Attestations	
1	adjectief (+*e*)_fout	flexion of the adjective	34	2
2	bezittelijk vnw (+*e*)_fout	flexion of the possessive pronoun	89	1
3	diminutief -*ke*	diminutive	93	13
4	*ikke*	personal pronoun *I*	5	1
5	object *u*	object *you*	189	8

Table 22. List of the morphological features that were used in case study 3

Feature 1:   **adjectief (+*e*)_fout**
Translation: wrong 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..'
Document n°:   SUD1.2

Feature 2:   **bezittelijk voornaamwoord (+*e*)_fout**
Translation: wrong 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?'
Document n°:   SUC4.1

Feature 3:   **diminutief -*ke***
Translation: diminutive
Corpus example: ***Pake**, voorzichtig.*
        '**Daddy**, be careful.'
Document n°:   SUF4.1

Feature 4:   *ikke*
Translation: flexion of the personal pronoun *I*
Corpus example: ***Ikke** Samson.*
        '**I** [want] Samson'
Document n°:   SULE1.2

Feature 5:   **object *u***
Translation: object *you*
Corpus example: *Ik ben ook niet verliefd op **u**.*
        'I am not in love with **you** either.'
Document n°:   SUF3.2

Feature	Colloquial Belgian Dutch	Translation	Attestations	
1	comparatief + *dan* + object	comparative	2	1
2	*durven* + inf	*to dare* + inf	2	1
3	*zijn ontslag geven*	*to resign*	3	2
4	*niet moeten*	*not have to*	5	5
5	vz + vz	preposition + preposition	6	1
6	*zijn*	possessive *zijn*	4	3
7	*onze/ons* + soortnaam/eigennaam	*our* + generic/proper name	5	5
8	*de* + eigennaam	*de* + proper name	11	1
9	*zet je erbij*	*have a seat*	1	1
10	aux + part + inf	position of the participle in the verbal end group	9	4

Table 23. List of the syntactic features that were used in case study 3

Feature 1:   **comparatief +** *dan* **+ object**

Translation: comparative

Corpus example: *Ik heb meer dan* **jou.**

     'I have more than **you**.'

Document n°:   SUCT1.1

Feature 2:   *durven* **+ inf**

Translation: *to dare* + infinitive

Corpus example: *We* **durven** *uw event niet* **organiseren**, *we zijn nog niet klaar.*

     'We don't **dare to organize** your event, we are not ready yet'

Document n°:   SUF3.2

Feature 3:   *zijn ontslag geven*

Translation: *to resign*

Corpus example: *Ja, daarom heb ik mijn ontslag gegeven.*

     'Yes, that is why I have resigned.'

Document n°:   SUF3.1

Feature 4:   *niet moeten*

Translation: *not have to*

Corpus example: *Je* **moet niet** *bang zijn.*

     'You **don't have to** be afraid'.

Document n°:   SUC1.1

Feature 5:   vz + vz

Translation: preposition + preposition (+*en*)

Corpus example:	*..en probeert zich niet **naar voor** te schuiven als het machtsblok.*
	'..and doesn't try to move **forwards** as power block.'
Document n°:	SUD1.1

Feature 6:	*zijn*
Translation:	possessive *zijn*
Corpus example:	*Nathalie, mag je iemand anders **zijn** kind straffen?*
	'Nathalie, is one allowed to punish someone else**'s** child?'
Document n°:	SUD4.2

Feature 7:	*onze/ons* + eigennaam/soortnaam
Translation:	*our* + proper name/generic name
Corpus example:	*Ja, ik ben hier met mijn man en **onze** Harry en Babette.*
	'Yes, I am here with my husband and **Harry** and Babette.'
Document n°:	SUF4.1

Feature 8:	*de* + eigennaam
Translation:	*the* + proper name
Corpus example:	*Dit is **de** Frans.*
	'This is **Frans.**'
Document n°:	SULE4.1

Feature 9:	*zet je erbij*
Translation:	*have a seat*
Corpus example:	*Allee Hélène, **zet je erbij.***
	'Come on, Hélène, **have a seat.**'
Document n°:	SUF3.1

Feature 10:	**aux + part + inf**
Translation	position of the participle in the verbal end group
Corpus example:	*Die risicopremie zal **moeten betaald worden.***
	'This insurance premium **must be paid.**
Document n°:	SUD1.2

Feature	Colloquial Belgian Dutch	Translation	Attestations	
			Spoken	Subtitles
Morphological features				
1	verbuiging lidwoord	flexion article	227	0
2	verbuiging aanwijzend vnw	flexion demonstrative pronoun	63	0
3	*ge/gij*	personal pronoun *you*	461	0
4	1ste persoon enkelvoud + *n*	deviant conjugation 1st p.s.	23	0
5	2^e persoon enkelvoud	deviant conjugation 2nd p.s.	86	0
Syntactic features				
6	*van/voor* + beknopte bijzin	reduced clause	6	0
7	dubbele negatie	double negative	7	0
8	redundant *dat*	redundant *that*	79	0
9	subjectsreduplicatie	reduplication of the subject	117	0

Table 24. List of additional CBD features that were used in case study 3

Feature 1:   *verbuiging lidwoord: den, ne(n)*
Translation: flexion of the article
Corpus example: *Hij zal **ne** fantastische papa zijn.*
      'He will be **a** great dad.'
Document n°:   SUF4.1

Feature 2:   *verbuiging aanwijzend vnw: diene(n), dieje(n), dezen*
Translation: flexion demonstrative pronoun
Corpus example: *Fantastisch nummer trouwens, **diene** Hold Back The River.*
      'Great song, **that** Hold Back The River.'
Document n°:   SULE2.1

Feature 3:   *persoonlijk voornaamwoord ge/gij*
Translation: personal pronoun *you*
Corpus example: *Ik weet nie of **ge** mij kunt horen, oma.*
      'I don't know whether **you** can hear me, grandma.'
Document n°:   SUCT1.1

Feature 4:   1ste persoon enkelvoud + *n*
Translation: deviant conjugation 1st person singular
Corpus example: *Ik **zen** gelukkig, hè.*
      'Well, I **am** happy.'
Document n°:   SUC1.1

Feature 5:   2e persoon enkelvoud
Translation: deviant conjugation 2nd person singular

Corpus example:	*Gij **zijt** 't probleem.*
	'You **are** the problem.'
Document n°:	SUF3.2

Feature 6:	***van/voor** + beknopte bijzin*
Translation:	reduced clause
Corpus example:	*Deze sla is te goed **voor te versnijden.***
	'This lettuce is too good **for cutting**.'
Document n°:	SULE4.1

Feature 7:	**dubbele negatie**
Translation:	double negative
Corpus example:	*Da had ik **nooit nie** gedacht da ik da ging winnen.*
	'I **never** thought I would win this game.'
Document n°:	SULE1.1

Feature 8:	**redundant *dat***
Translation:	redundant *that*
Corpus example:	*Wa **da** gij voor de kinderen doet, dat is onbetaalbaar.*
	'The things you do for the children cannot be repaid.'
Document n°:	SUF4.1

Feature 9:	**subjectsreduplicatie**
Translation:	reduplication of the subject
Corpus example:	*Ja, **ge** moogt **gij** de creativiteit van de jeugd ook nie aan banden leggen, hè.*
	'**You** must not curb the creativity of youth.'
Document n°:	SUC2.1

6.3   Results and discussion

6.3.1   Quantitative analysis

In this section, the results of the 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 testing the first and second hypothesis. Next, we will focus on the influence of the program genre on the subtitlers' linguistic choices in order to verify the third hypothesis.

Contrary to case studies 1 and 2, we will not apply profile-based correspondence analysis, since we do not want to determine the frequency distributions of BSD words and constructions compared to their CBD counterparts in different subtitling contexts. Instead, it will be calculated how often Flemish subtitlers reproduce the spoken colloquialisms from the original source text rather than converting these colloquial variants into BSD in order to investigate to what extent the subtitlers' linguistic choices are influenced by the original footage of the television program. Therefore, the ratio between CBD and BSD in the subtitle corpus in relation to the use of CBD in the spoken corpus is calculated manually and represented by means of diagrams. First, the CBD variants in the spoken corpus are counted. Next, both the colloquial variants and the standard variants in the corresponding subtitles are counted. These absolute scores of CBD variants and BSD variants in the subtitle corpus are then divided by the total number of CBD variants in the spoken corpus. This results in a relative frequency score which indicates how often subtitlers reproduce the colloquial variants from the spoken source text in the corresponding subtitles or how often they translate those colloquial variants into standard language. This calculation is made for the three feature sets (cf. Table 21 to 23) 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 are made for each program genre separately in order to verify whether the program genre influences the subtitlers' linguistic choices. Because of the small size of the corpus materials, the frequency scores of the individual language variants are not further analyzed.

6.3.1.1   General observations

A first look into our corpus data demonstrated that subtitlers use the colloquialisms only if the spoken source text contains this colloquial feature. In other words, subtitlers never convert a BSD variant in the original speech into a colloquial variant in the subtitles. However, this observation only counts if the original speech is Belgian-Dutch. Since this case study exclusively focuses on intralingual subtitling, we do not know, for instance, whether subtitlers add a Belgian-Dutch colloquialism to the interlingual subtitles of a non-standard foreign dialogue (e.g. whether the colloquial variant *cop* or *copper* 'policeman' is translated by the colloquial variant *flik*).

In the following analyses, we will investigate which Belgian-Dutch colloquialisms are used by the subtitlers. By doing so, we will not only be able to verify how often the spoken CBD variants are reproduced in the subtitles, but we will also reveal whether there are differences between lexical, morphological, and syntactic features.
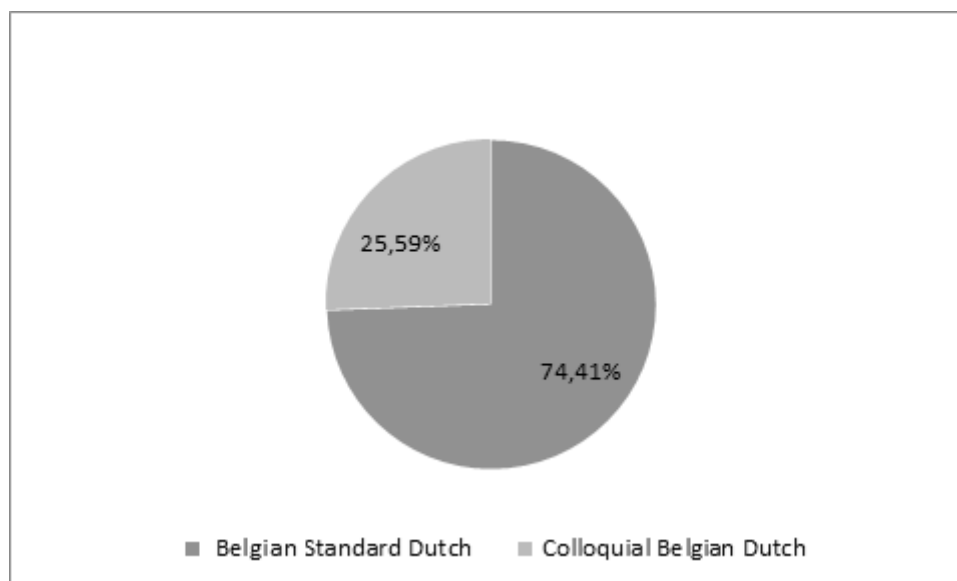
Figure 22. The ratio between CBD lexemes and morphosyntactic constructions (light grey) and BSD lexemes and morphosyntactic constructions (dark grey) in the subtitle data of the specialized parallel corpus

Figure 22 shows the ratio between the CBD and the BSD variants that were used in the subtitles. The relative frequency of CBD is marked in light grey, whereas the relative frequency of BSD is marked in dark grey. As expected on the basis of the previous case studies, it can be observed that subtitlers make use of CBD variants in intralingual subtitles on Flemish television. Furthermore, the diagram shows that only 25.59% of the spoken CBD variants are reproduced in the subtitles, whereas 74.41% is translated into BSD. In other words, subtitlers tend to avoid the use of colloquial features to a large extent, which confirms our first hypothesis.

Figure 23. (from left to right) The relative number of lexical, morphological, and syntactic CBD features that were reproduced (light grey) or translated into BSD (dark grey) in the subtitle data of the specialized parallel corpus

In Figure 23, the ratio between CBD and BSD variants is visualized for each feature set separately. 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, thus confirming the second hypothesis. 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 does not come as a surprise. Unlike ten years ago, VRT promotes in its new version of the subtitling style guide (VRT 2009) the reproduction of colloquial lexicon to retain the authenticity of the television program and the characters. Furthermore, the number of colloquial attestations (cf. Section 6.2) suggests that these new guidelines are partially based on the variants' frequency of use: lexemes are overall less frequently used than morphological and syntactic features. By limiting the reproduction of CBD to colloquial lexicon, the subtitles remain free from tussentaal to a large extent. If, on the other hand, VRT opted for the reproduction of, for instance, the colloquial variant of the personal pronoun *ge/gij*, the subtitles would be full of non-standard language.

Nevertheless, the diagram shows that subtitlers in a significant number of cases (36.43%) translate the CBD lexemes into a BSD alternative. An in-depth analysis of the data set reveals that these results are largely influenced by the disproportion of three individual features: the CBD lexemes *efkes* ('just'), *gelijk* ('like'), and *seffens* ('later') 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. 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 the 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 similar 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 (cf. Figure 24). As such, each lexeme was assigned an equal weight, regardless of its frequency in the subtitles.

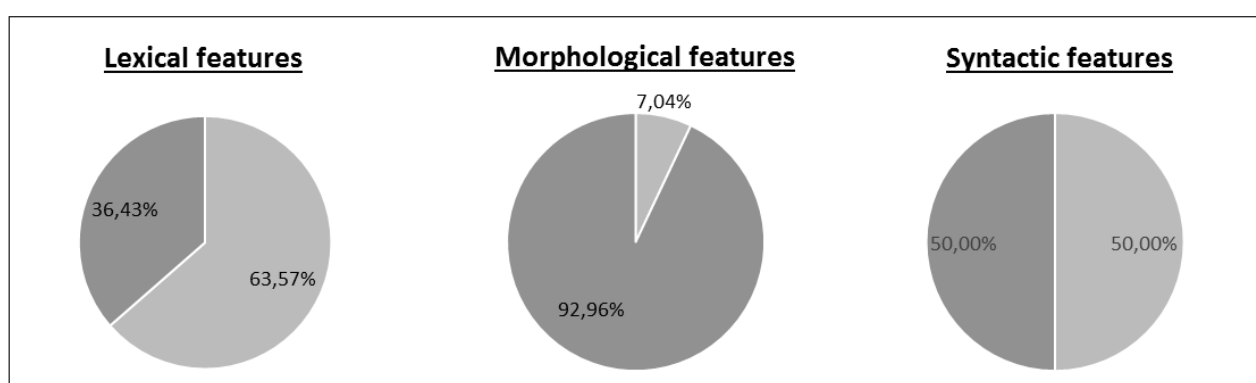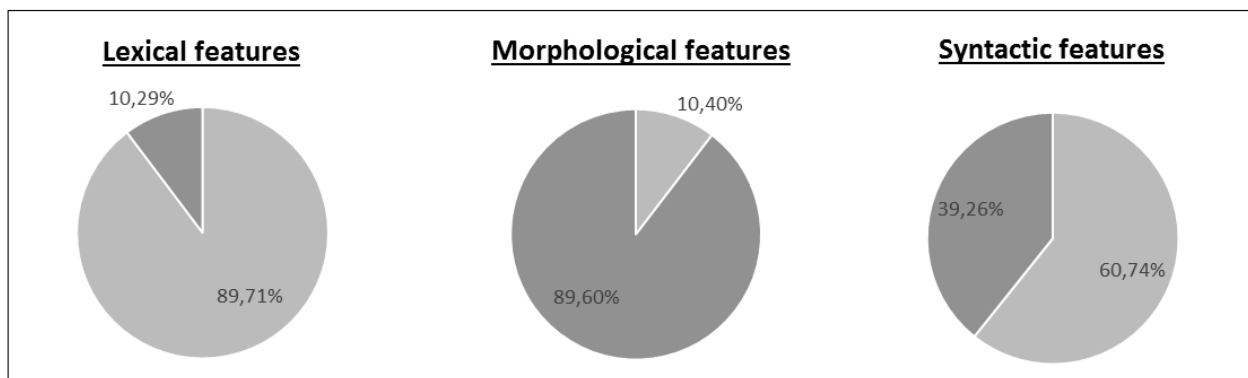Figure 24. Normalized frequencies of the relative number of lexical, morphological, and syntactic CBD features that were reproduced (light grey) or translated into BSD (dark grey) in the subtitle data of the specialized parallel corpus

The resulting diagram shows that subtitlers generally reproduce the CBD lexemes in the subtitles (89.71%), whereas only 10.29% of the lexical colloquialisms are converted into a BSD alternative. In other words, the calculation of the normalized frequencies of the lexemes confirms that the results in Figure 23 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, their position depending on how often they were reproduced or converted into BSD in the subtitles.

Figure 25. Continuum that visualizes the reproduction/conversion of the lexical features in the subtitle data of the specialized parallel corpus

In Figure 25, *seffens*, *gelijk*, and *efkes* are situated at the right side of the continuum, since these lexemes are generally converted into BSD in the subtitles. *Buizen* ('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 in Van Dale dictionary, subtitlers seem to handle them differently. 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 the study of Lybaert (2014b), it was 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 6.3.2, we will discuss this salience effect, together with other reasons for the frequent conversion of *seffens*, *gelijk*, and *efkes*.

For the morphological features, Figure 23 shows that the subtitlers copied only 25 of the 355 spoken CBD morphemes (7.04%) to the subtitles, whereas 330 spoken CBD features (92.96%) were replaced by a BSD alternative. In accordance with the outcome of the lexical features, the results for the morphological colloquialisms also match our expectations. In its style guide, VRT prescribes that morphosyntactic tussentaal items must be corrected, so it does not come as a surprise that subtitlers generally convert these CBD morphemes into BSD variants. However, in 7.04% of the cases, the colloquial morphemes are reproduced by the subtitlers. To eliminate a potential disproportion between the individual variants, we also calculated the normalized frequencies for the morphological set. The resulting diagram in Figure 24 shows, however, that this calculation does not have a great influence on the results. In addition, the continuum in Figure 26 visualizes the occurrence of the individual CBD morphemes in the subtitles.

Figure 26. Continuum that visualizes the reproduction/conversion of the morphological features in the subtitle data of the specialized parallel corpus

This profound look into our data indicates that out of the 25 reproduced CBD morphemes especially the diminutive *-ke* (*n* = 13) and the informal object *u* (*n* = 8) 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, since Lybaert's (2014b) 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 major part of the CBD morphemes in the spoken source text is converted into BSD, the subtitlers 13 times opt for the reproduction of the diminutive -*ke* and 8 times for the reproduction of the informal object *u*, 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 6.3.2, we will discuss their comments in detail.

Finally, Figure 23 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 relative frequency of CBD constructions in the subtitles even increases to 60.74%. 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 must be corrected. On first thoughts, this remarkable observation could be explained by the fact that the syntactic domain is less salient than the lexical domain (Lybaert 2014b). According to Van Bree (2000), this can be attributed to the abstract nature of the syntactic constructions: unlike the lexicon, this domain is characterized by abstract rules, which makes it more automated or less concrete, so language users unconsciously use these features. However, when we put the syntactic features with at least five attestations along the continuum below, it can be seen that subtitlers deal differently with the individual variants.

CBD IN SUBTITLES   **BSD IN SUBTITLES**

niet moeten   *prep + prep*

possessive zijn   *de + proper name*

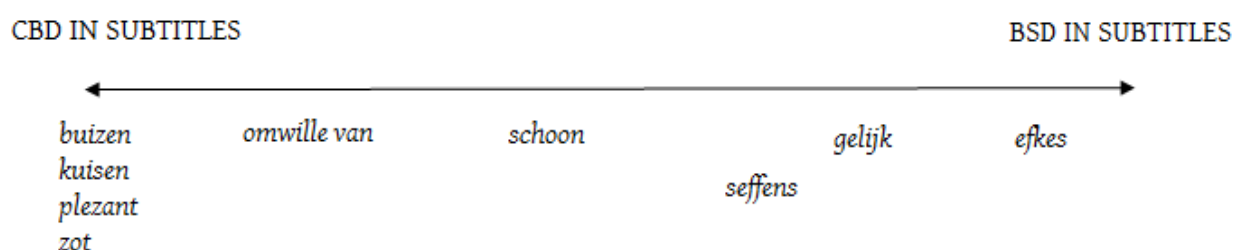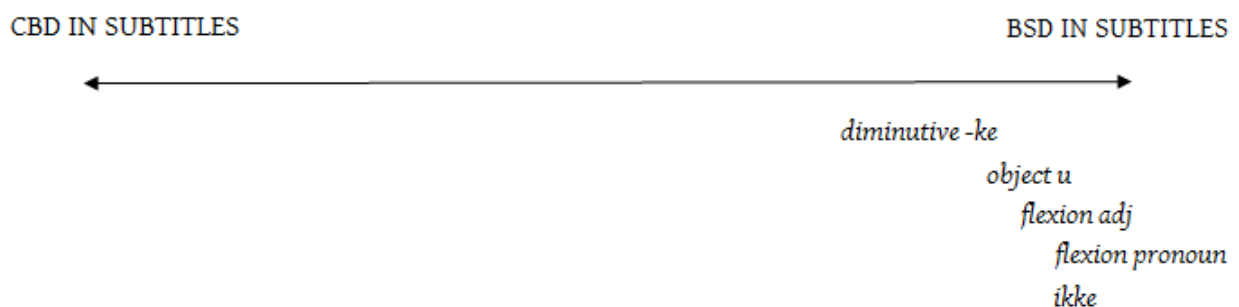*ons/onze + generic/proper name*   *aux + part + inf*

Figure 27. Continuum that visualizes the reproduction/conversion of the syntactic features in the subtitle data of the specialized parallel corpus

The continuum demonstrates that subtitlers frequently transfer three colloquial features in particular: *niet moeten* ('not have to'), the possessive *zijn*, and *ons/onze* ('our') + generic/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* + generic/proper name can be explained by the subtitling guidelines of the public broadcaster. In its style guide, VRT prescribes that the flexion of pronouns is not allowed in the subtitles, with the exception of the construction *ons/onze* + generic/proper name. Other syntactic colloquialisms, however, 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 reasons why subtitlers opt for the reproduction of these colloquialisms in particular. In Section 6.3.2, 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 opt for the reproduction of CBD morphemes (7.04%) and constructions (50.00%), although tussentaal morphosyntax should always be corrected according to VRT's style guide. As a consequence, the question arises by which factors the subtitlers are driven to reproduce these morphological and syntactic colloquialisms. In case studies 1 and 2, we have already 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 this new subtitle corpus in order to find out whether the language choices made in subtitles produced for Flemish television differ according to the program genre.

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

In case study 2, it was 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*). Since the present case study is based on a new parallel corpus, containing more recent subtitling material, we want to verify whether the program genre still influences the subtitlers' linguistic choices. Furthermore, we want to present the results of the quantitative analyses to the subtitlers and ask their vision on this outcome (cf. Section 6.3.2). In this section, it will therefore be indicated in which genres the subtitlers

reproduce the Belgian-Dutch colloquialisms from the spoken source text and whether their choices differ depending on lexical, morphological, and syntactic variants.
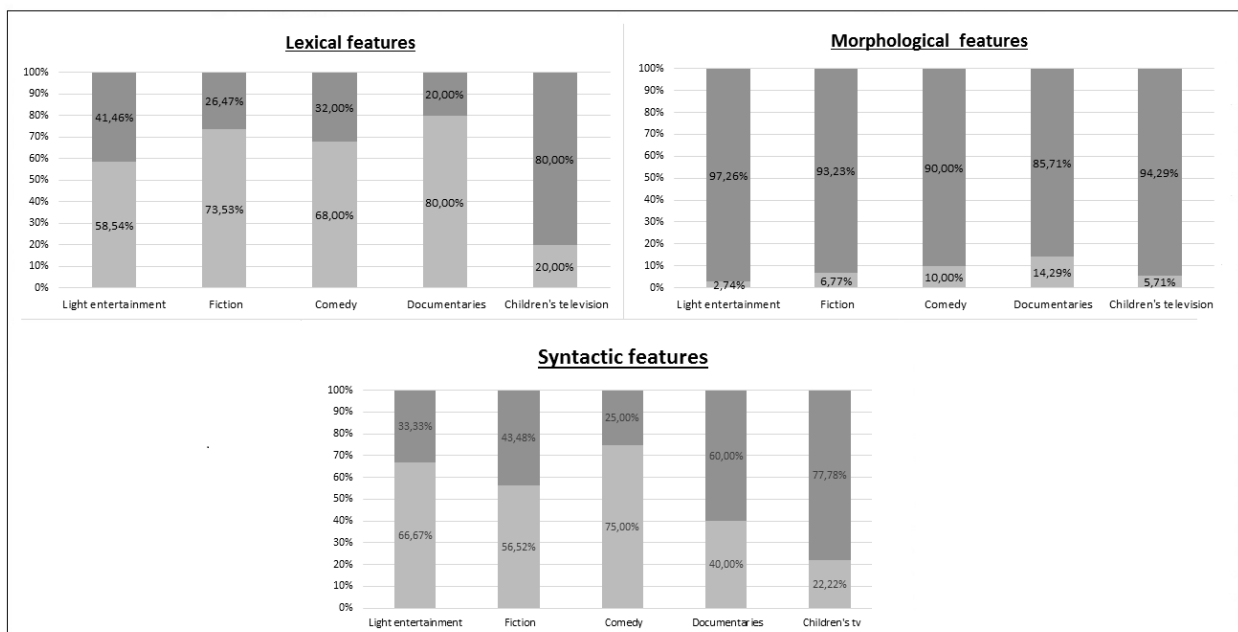
Figure 28. The the relative number of lexical, morphological, and syntactic CBD features that were reproduced (light grey) or translated into BSD (dark grey) in five program genres in the subtitle data of the specialized parallel corpus

In Figure 28, the ratio 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 syntactic constructions from 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) yields a distorted picture. Table 26 gives an overview of the distribution of BSD and CBD features in the subtitles of the five program genres, with *n* representing the absolute number of attestations and *%* representing the relative number of attestations.

Data set	Label	Program genre									
		Light entertainment		Fiction		Comedy		Documentaries		Children' television	
		n	%	*n*	%	*n*	%	*n*	%	*n*	%
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

Table 25. Overview of the absolute and relative numbers of BSD and CBD attestations per feature set for the five program genres in case study 3

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* than in *documentaries*. In the diagram with the morphological features of Figure 28, 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* = 14), 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 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 greater amount of CBD lexemes and syntactic constructions in subtitles of *comedy*, and to a certain extent in those of *fiction* and *light entertainment*, confirm our third hypothesis, viz. that the number of CBD variants in the subtitles is higher in entertainment, infotainment and humor programs (vs. informative programs). The linguistic choices that are made in these genres can be explained by the general aim of these programs. The main objective of humor 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 for the sake of the authenticity of entertainment programs (Hendrickx 1998). As a result, CBD is

frequently spoken in these genres[25] and the subtitlers seem to reproduce these colloquial features in the subtitles to a large extent. By doing so, they retain the informal, spontaneous nature of the TV program and they avoid that the characters' speech corresponds too much to the written language of the printed script (Rosa 2001). Conversely, the educational role of *children's television* could explain why subtitlers translate the (sporadically spoken) 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 *Taalcharter*, the public broadcaster VRT emphasizes that the use of standard language is required in children's programming (Hendrickx 1998; 2012).

6.3.1.3 An additional analysis of some typical CBD features that did not appear in the subtitles

As mentioned in Section 3.2.3, one of the consequences of the methodology that we used to gather our data is that we only selected the colloquial features that appeared in the subtitle corpus. An underlying argument to adopt this selection procedure is that we are primarily interested in which features are used in the subtitles. As a consequence, other typical CBD elements were not included in the data set, because subtitlers translate these CBD elements into BSD. However, even though these elements did not occur in the subtitles, it is interesting to examine how often subtitlers convert them into standard language. To solve this shortcoming, we consulted existing lists of tussentaal ingredients and added a set of nine commonly used CBD features to our data. Table 25 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.

Feature	Colloquial Belgian Dutch	Attestations in spoken source text
	Morphological features	
1	flexion article: *den, ne(n)*	227
2	flexion demonstrative pronoun: e.g. *dieje(n), diene(n), dezen*	63
3	personal pronoun *ge/gij*	461
4	deviant conjugation 1st person singular (*+n*)	23
5	deviant conjugation 2nd person singular	86
	Syntactic features	
6	*van/voor* + reduced clause	6

---

[25] The number of spoken CBD attestations is 190 for *fiction*, 154 for *comedy*, and 120 for *light entertainment*, whereas in *documentaries* and *childrens' television*, these numbers are merely 29 and 54 respectively.

7	double negative	7
8	redundant *that*	79
9	reduplication of the subject	117

Table 26. Total number of attestations of typical CBD features in the spoken source text

In Table 26, three features immediately attract attention because of their high number of occurrence in the spoken corpus: the personal pronoun *ge/gij* (*n* = 461), the flexion of the article (*n* = 227), and the reduplication of the subject (*n* = 117). First, it does not come as a surprise that subtitlers do not reproduce these CBD features in the subtitles, since VRT's guidelines prescribe that morphosyntactic colloquialisms must be converted into standard language. Nevertheless, the analyses in Section 6.3.1.1 have shown that other morphological and syntactic colloquialisms do occur in the subtitles, although subtitlers are expected to translate them into BSD. Second, the total number of attestations of each of the spoken CBD features in Table 26 is significantly higher than the total number of attestations of the reproduced colloquialisms in Table 22 and 23. Nevertheless, all features in Table 26 were converted into BSD in the corresponding subtitles, presumably because their frequent occurrence would substantially enhance the colloquial character of the subtitles. This could explain why subtitlers tend to avoid these colloquial items. In the following section, it will be explored in depth why subtitlers opt either for the reproduction or for the conversion of the CBD features.

6.3.2 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, it has also become clear that the program genre influences the subtitlers' linguistic choices. Furthermore, we have seen that subtitlers do not consistently 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, who had been involved in the development of VRT's subtitling guidelines, 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. In the discussion, quotes of the interviewees have been translated.

6.3.2.1   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 *spreektaal* (colloquial) or *informeel* (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. *oom* ['uncle'] counts less characters than *nonkel* ['uncle']). 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 a compromise that 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. VRT does, however, not comply with this desire regarding grammatical constructions. Furthermore, 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'. In other words, the subtitlers' linguistic choices are driven by the program genre, which confirms our quantitative findings in case studies 2 and 3. This language behavior is, however, a choice of the subtitlers themselves, since these genre considerations are not mentioned in the subtitling style guide.

   With regard to morphosyntactic constructions, the subtitlers follow 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 are the 'critical viewers who favor the use of standard language'. 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 standard language. Although subtitler 2 does not always agree with these decisions, '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 'actual practice is merely an approach to the ideal'. 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. The subtitles are sent to a fellow-worker who removes typing mistakes, language errors, and other linguistic irregularities. 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.

6.3.2.2 Subtitlers' perceptions of the subtitling reality at VRT: some corpus examples

The quantitative results in Section 6.3.1.1 have shown that lexical colloquialisms are more frequently used in the subtitles of Flemish 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 costs 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 did that with **onion** too.

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 presented these corpus examples to the subtitlers, they declared that the five examples are completely in accordance with the subtitling guidelines. First, the colloquial diminutive *-ke* 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. As a result, 'it would be odd to read this *-ke* diminutive in the subtitles', so 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" colloquial', 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 (e.g. *[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 2007). In closed intralingual subtitling, however, 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 dialectical', 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* 'would you'). According to subtitler 2, these colloquial variants are never used in the subtitles, because they often cause grammatical issues. The personal pronoun *ge/gij*, for instance, requires a colloquial conjugation of the verb (e.g. *ge zijt* and not *ge bent* 'you are). The reproduction of this CBD variant of the verbal form (*zijt*) is barred by VRT's official subtitling policy. 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 *zijt* into standard language, viz. *je/jij bent* ('you are').

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

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. On the other hand, they indicate that there was a lot of discussion about *efkes* and *seffens* 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 issues, 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, but not always:

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**.

Subtitler 1 declares that *schoon* has a similar connotation as *efkes* and *seffens*: 'In the subtitlers' mind, this word sounds "very Flemish"', by which she means that it sounds too dialectical and VRT's subtitling policy does not allow the reproduction of dialect words in the subtitles. This results in 'a reflex to "clean up" the dialectal speech by translating *schoon* into standard language', even though Van Dale dictionary labels this lexeme as 'colloquial'. Subtitler 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 next time, and vice versa. Usually, 'errors' similar to example 9 are eliminated during the final editing.

Contrary to their tolerant attitude toward the use of CBD 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 are occasionally reproduced in the subtitles, and syntactic colloquialisms even quite often (viz. in 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 ought to be systematically replaced by the BSD *je/jou*-form, just like the subject *ge/gij* is translated into *je/jij.* However, it occurs on occasion that subtitlers unconsciously copy the *u*-form to the subtitles The subtitlers attribute this 'error' 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. Subtitler 2 admits that 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. In other words, subtitlers perceive the informal object *u* as a lowly salient feature. 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* played 'a pioneering role'.[26] Some characters in the soap are systematically addressed by

---

[26] *Thuis* ('Home') plays out the daily life of several upper, lower and working class families and is broadcast daily by VRT since 1995.

a diminutive that is almost considered as their real name (e.g. *Simonneke* 'Simonne', *make* 'mommy', *Lowieke* 'Lowie'), 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 6.3.1.1 not only pointed out that morphological colloquialisms are occasionally reproduced in the subtitles, but especially 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. 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. Acoording 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 mistakes 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 deviant position of the participle in the verbal end 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. Given the abovementioned findings, the subtitlers largely confirmed our assumptions of hypothesis 3.

6.4 Concluding remarks

By comparing the original speech in 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, we investigated to what extent the subtitlers reproduce the lexical, morphological, and syntactic colloquialisms of the original footage, 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. Both the quantitative and the qualitative analyses have revealed that subtitles are the result of a delicate process to find a linguistic balance between several, and at times contradictory, criteria. One the one hand, the subtitlers have to follow the official language policy that aims to guarantee the general intelligibility of the TV programs. On the other hand, the subtitlers have to take into consideration the authenticity of the program genre, and the needs and requirements of different audience groups who often have opposite expectations. Finally, technical criteria like the limited number of characters and pressure of time have an impact on the subtitlers' linguistic choices. In other words, the subtitlers' work cannot be explained 'just in terms of the strategies used, but also how they are used in interaction with the other elements of the audiovisual product and the specific parameters of a given time and space" (Pinto 2017: 17).

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 toward the use of colloquial lexemes in the subtitles. By doing so, the public broadcaster has worked out a compromise between the conflicting wishes of the audience. By reproducing the colloquial lexicon, VRT wants to meet at least partially the deaf and hard-of-hearing audience's wishes to repeat the spoken source text as literally as possible in the subtitles on the one hand. On the other hand, morphosyntactic contructions are converted into standard language to satisfy the viewers who take offence at linguistic errors. However, our results have demonstrated that within the morphological field, 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 non-salient, colloquial variant. For similar reasons of salience, some syntactic colloquialisms are frequently used in the subtitles. According to the subtitlers, the deviant 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. In other words,

certain morphological and syntactic colloquialisms seem to be highly entrenched into the subtitlers' cognitive language system and the use of these CBD variants 'has become a highly automated routine' (Schmid 2007: 118). Several studies have associated this cognitive entrenchment to word frequency (e.g. Bardovi-Harlig 1987; Bybee 2003; Schmid 2010). According to Langacker (1987: 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, the aforementioned studies presume that there is also a direct relation between the frequency of a variant and its salience: the more frequent a variant is used, the more salient it becomes, and vice versa (Halverson 2017). 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: 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.

Chapter 7
Conclusion

Research in the field of Corpus-Based Translation Studies has repeatedly demonstrated that translated texts contain more neutral expressions, more conventional and less creative language compared to their source texts or comparable non-translated texts, which is assumed to be related to a standardizing, norm-adhering trend. Nevertheless, recent studies have shown that this standardization tendency is just a tendency, and not a universal, as its occurrence depends on contextual parameters as *genre*, *source language*, and *target audience* (e.g. Delaere et al. 2012; Delaere and De Sutter 2013; De Sutter et al. 2012a; Kruger & van Rooy 2012). For the Dutch language area, for example, Delaere (2015) has demonstrated that in general, translators of 'regular' written genres more often opt for commonly accepted Standard Dutch words and constructions compared to writers of original texts (non-translators), but to which their lexical and grammatical choices conform to the Standard Dutch norm depends on extralinguistic factors such as *source language* and *register* or *genre*. Building on Delaere's (2015) research, the present dissertation has placed the study of the language used in Flemish subtitling at the center of its concerns. Subtitles are characterized by a colloquial input and are thus situated between oral and written language, which makes them a special type of translation. On the one hand, as subtitles are the result of the transformation of spoken language, with its typical colloquial and often non-standard features, into written text, subtitlers might be encouraged to use non-standard linguistic items (Díaz-Cintas 2010: 344-346; Karamitroglou 2000; Neves 2004). On the other hand, subtitles are also heavily edited translations, which might stimulate the use of standard language. This raises the question what kind of linguistic choices subtitlers make.

Furthermore, the specific language situation in Flanders makes this question all the more compelling. Even though the Flemish public broadcaster VRT has developed an official language policy which strongly clings to the use of BSD by television and radio hosts, several studies have suggested that in actual practice, CBD is frequently used in spoken public media (e.g. Prieels 2013; Van Hoof 2010; 2015). These language dynamics

are often attributed to a process of destandardization, which implies that 'the established standard language loses its position as the one and only "best language"' (Coupland & Kristiansen 2011: 28). Given the recent linguistic dynamics in spoken standards, not only in Flanders, but all over Europe (Odendaal 2014; Ghyselen et al. 2016), the question is raised whether these processes also affect written language. Nevertheless, the influence of these spoken language changes on the written language has barely been taken into consideration.

To fill this gap, the main goal of this dissertation was to investigate how subtitlers deal with the norm-related linguistic tension that exists between VRT's norm-adherent language policy on the one hand, and the specific Flemish language reality on the other hand. Given the wide distribution of CBD, not only in informal spoken contexts, but also in the spoken language of public institutions, our research verified whether this colloquial variety also infiltrates into the written language (viz. subtitles) of the Flemish public broadcaster, which is traditionally considered to have an authoritative function with regard to language use. More concretely, it was quantitatively investigated to what extent subtitlers in Flanders choose CBD words and constructions instead of BSD words and constructions, and how these choices can be compared to the linguistic choices made in Dutch translations and original Dutch texts. In addition, it was verified which contextual factors influence the use of colloquial variants vs. standard variants by applying multivariate statistical techniques. Contextual factors that were included in these analyses are *source language* (English vs. Belgian Dutch vs. Netherlandic Dutch), *program genre* (e.g. fiction vs. documentaries), *program purpose* (e.g. informing vs. entertaining), *target audience* (adults vs. children), and *cast* (e.g. actors vs. non-actors). In order to gain insight in the motivation behind Flemish subtitlers' linguistic choices, an additional qualitative analysis was performed by means of semi-structured interviews, evaluation reports, and observational data that were collected at the subtitling department of the Flemish public broadcaster. By doing so, the present study will offer an insight into the (de)standardization tendencies in written subtitles on Flemish television.

The first theoretical chapter of this dissertation gave an overview of the existing literature on audiovisual translation. Furthermore, we zoomed in on some norm-related concepts to provide the reader with some background information against which this research is situated. In addition, we outlined the Dutch standardization process, together with its impact on the current language situation in Flanders and the language policy of the Flemish public broadcaster. Standard languages currently seem to be under pressure all across Europe and broadcasting media are believed to play a central role in the distribution of non-standard, colloquial language. In Flanders, the increasing use of CBD on television has even caused a shift in the language policy of the public broadcaster, which initially rejected CBD. In recent years, however, VRT has become more tolerant toward the use of other varieties than BSD, especially in fiction and comedy programs.

Furthermore (and as a consequence), several studies have indicated that spoken language on Flemish television nowadays shows a lot of variation, even in programs that actually require the use of BSD (e.g. Prieels 2013; see also the evaluation reports of the project *Taalhantering* of KU Leuven & VRT 1999). In this dissertation, we aimed to investigate the extent to which the written subtitles of the Flemish public broadcaster conform to the BSD norm.

In the methodological chapter, we presented the corpus materials on which the three case studies were based. We also discussed the selection process of the linguistic variables as well as the annotation procedures, the program genre classification, and the statistical techniques that were used in the study of VRT television subtitles. We introduced profile-based correspondence analysis, a statistical technique which allows the visual exploration of our data set in order to find associations and patterns in the data. These analyses measured the linguistic choices of the subtitlers and visualized them in a two-dimensional plot. In addition, we introduced the qualitative analysis that was performed to explain the results obtained in the quantitative study.

Finally, chapters 4 to 6 presented the three case studies that were carried out in this dissertation. The first case study (Chapter 4) compared the linguistic choices made in written translations and non-translations to the linguistic choices made in interlingual and intralingual subtitles. Furthermore, the study investigated which contextual parameters influenced the language use in the subtitles. Case study 2 (Chapter 5) verified whether the subtitles contain more colloquial lexemes than colloquial grammatical constructions, and which contextual parameters influenced these linguistic choices. Finally, case study 3 (Chapter 6) addressed the question to what extent Flemish subtitlers reproduce Belgian-Dutch colloquialisms from the spoken source text and how they explain these linguistic choices.

7.1 Empirical findings: revisiting the research questions

By setting up a multifactorial study, we investigated which contextual factors influenced the subtitlers' norm-related linguistic choices and how these factors are related to each other. Furthermore, a qualitative analysis based on semi-structured interviews, evaluation reports, and observational data that were collected at the VRT's subtitling department contributed to improve our insight into subtitlers' practices and attitudes toward the VRT's language policy and the language reality in Flanders. As a result, we provided answers to five research questions that were formulated at the beginning of this dissertation:

1. Do Belgian-Dutch subtitles contain more or fewer linguistic features typical for Belgian Standard Dutch than for Colloquial Belgian Dutch in comparison with other translated and non-translated written genres?
2. To what extent do Flemish subtitlers reproduce the spoken Belgian-Dutch colloquialisms in the subtitles or do they even add colloquialisms to the subtitles?
3. To what extent do Flemish subtitlers adhere to the language policy of the Flemish public broadcaster, regarding the use of colloquial lexemes and colloquial grammatical constructions?
4. Which contextual factors (e.g. *source language*, *program genre*, *target audience*) have an influence on subtitlers' linguistic choices?
5. How do the subtitlers themselves explain their linguistic choices?

In case study 1, we investigated linguistic norm adherence in written translations and non-translations, and we compared this to the linguistic choices of Flemish subtitlers, thus answering the **first research question.** Building on two large corpora of Belgian-Dutch written and audiovisual translation, we investigated the dispersion of BSD vs. CBD in subtitling and in other written translations and non-translations in order to demonstrate which text types conformed to the BSD norm and which ones exhibited more Belgian-Dutch colloquialisms. The results of this first case study have revealed that translators were more norm-adherent than non-translators. Furthermore, it was shown that there was a considerable difference in language use between audiovisual and written translations, as the subtitles contained significantly more CBD variants compared to regular written translations. In-depth analyses pointed out that linguistic choices in these subtitles were mainly determined by *source language* and by *speaker type* (cf. research question 4).

The **second research question** was highlighted in case study 3, which compared the original speech in twenty Flemish television programs to the corresponding intralingual subtitles in order to investigate the extent to which subtitlers reproduce the colloquial variants of the original footage. The results have demonstrated that subtitlers used these colloquialisms only if the spoken source text contained them as well. In other words, subtitlers did not add CBD features to the subtitles if they did not appear in the original speech. Furthermore, it was shown that subtitlers tended to avoid the use of colloquial language, since the majority of the spoken colloquialisms was translated into BSD in the subtitles. Nevertheless, in-depth analyses indicated that subtitlers' linguistic choices largely depended on the *program genre* (cf. research question 4), and that their language choices differed with regard to lexical, morphological or syntactic features (cf. research question 3).

Case studies 2 and 3 both answered the **third research question.** The results of case study 2 have shown that non-standard lexemes were less frequently used by the subtitlers

than non-standard grammatical constructions. In other words, Flemish subtitlers were generally more norm-adhering when it came to lexemes than to grammatical constructions. On the contrary, the results of case study 3 have revealed that CBD lexemes were frequently reproduced in subtitles on Flemish television, whereas colloquial morphological items were generally converted into BSD. Syntactic colloquialisms took up a middle position. These contradictory results can be attributed to the use of different corpora in both case studies as well as to the modified subtitling guidelines of the Flemish public broadcaster. Not only is there a time difference of more than ten years between the subtitle materials of case study 2 and case study 3, VRT also recently became more tolerant toward language variation in television subtitling compared to ten years ago, especially when it comes to the use of colloquial lexicon. Furthermore, it was demonstrated that subtitlers neither consistently reproduced the colloquial lexemes, nor did they convert every morphological or syntactic colloquialism into BSD. In general, we explained these results in terms of salience and cognitive entrenchment. On the one hand, the most obvious explanation offered for the results in case study 2 was that CBD lexical features are more salient, and thus significantly more often perceived by language users, than CBD syntactic features are. As a consequence, subtitlers tended to avoid non-standard lexicon. On the other hand, the unconscious reproduction of certain morphological and syntactic colloquialisms in case study 3 was explained by their frequent use in everyday language. Since these commonly used colloquial variants are highly entrenched into the subtitlers' cognitive language system, they often did not notice them. In Section 7.2, we will elaborate on these tentative explanations.

In each case study, we investigated the influence of several contextual parameters on the subtitlers' linguistic choices, thus answering the **fourth research question**. In case study 1, the linguistic choices in the subtitles were largely determined by the *source language* and by the *speaker type*. On the one hand, the frequency of CBD variants increased significantly when the source language of the television program was Belgian Dutch (compared to English or Netherlandic Dutch). In addition, the number of colloquial variants also increased significantly when the subtitled speech was produced by an actor or interviewee. The contextual parameter *program genre*, however, did not yield significant results, which we attributed to the rather general division consisting of two genre categories that we adopted in our research design.

As a consequence, a more fine-grained genre classification was applied in case study 2. These results demonstrated that the program genre caused norm-related differences in the subtitle corpus, as subtitlers were more norm-adhering when subtitling informative programs (*documentaries*) and programs with an educational role (*children's television*), whereas the number of CBD lexemes and constructions increased significantly in the subtitles of humorous (*comedy*) and entertaining programs (*fiction* and *light entertainment*). Furthermore, the subtitles of English or Netherlandic-Dutch spoken television programs contained generally more standard language than programs of which the source language

was Belgian Dutch. The most obvious explanation offered for these results is that subtitlers transferred the BSD (*documentaries* and *children's television*) and CBD (*fiction*, *comedy*, and *light entertainment*) variants that were used in the spoken source text directly to the subtitles, thereby maintaining not only the 'Belgian atmosphere', but also the entertaining or comic effect that was created in the original footage. In case study 3, the language use in the original, spoken television fragments was analyzed and compared to the corresponding subtitles to substantiate these assumptions.

The results of the third case study confirmed that the program genre influenced the linguistic choices that were made in the subtitles. Whereas the Belgian-Dutch colloquialisms were generally translated into standard language in the subtitles of informative genres and programs intended for children, the number of colloquial features increased significantly in the subtitles of humorous and entertainment programs to reproduce the spontaneous or comic effect of the original footage.

In sum, the results of these case studies emphasized the importance of contextual factors that should be taken into account when analyzing subtitlers' linguistic choices, as our research indicated that these language choices were largely influenced by the extralinguistic context.

In order to clarify the obtained findings of the quantitative analyses in case study 3, we interviewed the head of the subtitling department and two subtitlers at VRT. By doing so, we aimed to get a better insight into their ideas and opinions regarding the guidelines they have to follow in relation to the existence of several formal and informal language varieties in Flanders. In addition, we asked for their comments on some corpus examples. This qualitative approach gave an answer to the **final research question.** During the interviews, the subtitlers repeatedly declared that they attach great value to the normative advices of the Dutch language authorities (i.e. Van Dale dictionary, VRT's Language Charter, and *Taaladvies*) on which the subtitling guidelines of the public broadcaster are based. Furthermore, the interviews revealed that many decisions concerning the use of CBD in the subtitles were made in function of the audience. On the one hand, the deaf and hard-of-hearing want the subtitlers to render the spoken source text as literally as possible, whereas, on the other hand, the subtitlers were afraid that a lot of television viewers would complain if the subtitles contained grammatical mistakes. As a result, subtitlers continuously struggle to find a balance between the official language policy that aims to guarantee the general intelligibility of the TV programs and the needs and requirements of different audience groups who often have opposite expectations. Finally, technical criteria like the limited number of characters and pressure of time had an impact on the subtitlers' linguistic choices.

7.2   Theoretical implications

We set out this study under the assumption that Flemish television subtitling contains more CBD than written Dutch translations. One of the underlying arguments was the colloquial input, which positions subtitles between oral and written language. The results of our research have demonstrated that subtitlers indeed use more CBD words and constructions than translators (cf. case study 1) and that the use of CBD in the subtitles is largely caused by the interference of the Flemish source text (cf. case study 3). In other words, subtitlers transfer the spoken Belgian-Dutch colloquialisms to the subtitles. The reproduction of colloquial lexemes is a conscious strategy, as VRT promotes in its subtitling guidelines the use of colloquial lexicon to retain the authenticity of the television program and the characters. The reproduction of morphosyntactic colloquialisms, however, is not allowed by VRT's subtitling policy. Nevertheless, our research has shown that subtitlers neither consistently reproduce the colloquial lexemes, nor do they convert every morphological or syntactic colloquialism into BSD. In general, we explained these results in terms of salience and cognitive entrenchment.

Although CBD grammatical contructions are largely converted into standard language, certain morphological and syntactic colloquialisms are occasionally used in the subtitles. Some features (e.g. the informal object *u*, the deviant position of the participle in the verbal end group, and the construction *niet moeten*) seemed to be highly entrenched into the subtitlers' cognitive language system, because of their frequent occurrence in everyday language. As a result, the use of these CBD variants 'has become a highly automated routine' (Schmid 2007: 118; Van Bree 2000; see also Bardovi-Harlig 1987; Bybee 2003; Langacker 1987; Schmid 2010 for the relation between cognitive entrenchment and word frequency). In addition, our study has shown that subtitlers find it difficult to recognize these commonly used and deeply entrenched language features. As a consequence, subtitlers often unconsciously reproduce these grammatical colloquialisms, because they are not always aware of these non-salient, colloquial variants.

Nevertheless, our research has revealed that salience also works in the opposite direction, thereby fulfilling an important role in the style-shifting process (Vandekerckhove & Ghyselen 2017). Although the *-ke* dimunitive, for instance, is considered a highly salient CBD feature, and should thus easily be detected, subtitlers regularly reproduce this morphological colloquialism in the subtitles of fiction programs. In this program genre, the reproduction of the *-ke* diminutive attributes to the characterization of certain characters. Precisely because the *-ke* diminutive is a salient colloquialism, subtitlers use this feature to typify a character and to retain the authenticity of the TV program. In other words, the CBD features 'function as a mimetic resource of which the author [i.e. the subtitler, my addition] takes advantage for the

indirect depiction of the characters, the interpersonal relations established between them and the discursive situations' (Pinto 2017: 4; see also Lippi-Green 1997; Hodson 2014). In other program genres, such as documentaries, the reproduction of the *-ke* diminutive is not functional and due to its salience, it is easily noticed and eliminated by the subtitlers.

Next to the functional role that is attributed to the use or omission of salient language features, the results of this study have revealed that salience is perceived differently, according to the linguistic knowledge of the language users (cf. Vandekerckhove & Ghyselen 2017 for the influence of the informants' regional origin on salience perception). CBD features are noticed more easily by subtitlers, who can be considered language professionals, than by regular language users, whose metalinguistic knowledge of the official language norm can generally be considered more limited. During their training, subtitlers have been made alert to Belgian-Dutch colloquialisms that should be avoided in professional contexts. As a result, it can be assumed that every colloquial lexeme is equally salient to them. Nevertheless, subtitlers assign different levels of colloquiality to the individual colloquialisms and they try to assess which features will be perceived as more salient than other by the television viewers. Although they are expected to reproduce the CBD lexemes in the subtitles, subtitlers avoid colloquial lexemes that run the risk of being evaluated as too dialectal by the audience, thereby going against VRT's subtitling guidelines.

In general, it can be confirmed that nowadays CBD or tussentaal is no longer exclusively used in the spoken registers, since it also occurs in (written) subtitles. As a consequence, the question arises whether this will cause 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 dissertation, 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 to a large extent. Especially in more formal and more informative written contexts, the standard language ideal is still very much alive among Flemish subtitlers and BSD continues to be the outstanding variety to be used. According Jaspers & Van Hoof (2015: 35), a quantitative increase of non-standard, colloquial language use in the public domain, and especially on television, should not be seen as 'an undiluted sign of the dwindling hold of standardization on the public mind':

> A quantitative increase does not by itself change the conditions within which these quantities are produced [...]. To be sure, while media allows for much more linguistic diversity in drama and comedy, shows, and other entertainment, the

more "important" and authoritative genres (such as the news) are still the exclusive domain of the standard variety.

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: 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.

7.3   Limitations of the study and directions for further research

Beside the empirical findings and theoretical implications described in the previous sections, this research also had a number of limitations, which might be overcome in future work.

First, the visualization of the associations between the variables and the contextual factors in a two-dimensional plot makes profile-based correspondence analysis highly suitable for an exploratory analysis to find patterns in our data. Honesty compels us to mention, however, that correspondence analysis is merely a technique that allows us to observe the overall effects per factor, but the visualization of the exact causal relations goes beyond the capacity of this technique. Although the results of our analyses yielded a more fine-grained insight into the linguistic characteristics of Flemish television subtitles, additional research needs to be done to identify the causes for the observed trends. By carrying out, for instance, an explanatory analysis that includes all the

available metadata, it could be verified which factors cause the observed tendencies. Ideally, both our exploratory and an explanatory product-based study should be completed with an extensive process-based analysis to reveal when and why certain language choices are made by the subtitlers.

Second, the corpus-based approach has restricted our research to the study of the subtitles as a product and not as a process. As a consequence, we were not able to determine where exactly in the subtitling process a certain linguistic choice was made: was the norm-related variant chosen by the subtitler himself or by the editor during the final editing process? Furthermore, the metadata of the consulted corpora were often limited or incomplete, which also prevented us from generating conclusions with regard to the language choices made by an individual subtitler. For that reason, we would like to make an appeal for subtitle corpora enriched with detailed, accurate, and complete metadata.

In addition to corpora with more complete metadata, there is also a need for subtitling corpora that include the original spoken source text. The absence of the original audio text prevented us more than once from verifying the impact of the spoken source text on the subtitlers' linguistic choices. Furthermore, the consulted subtitle corpus dates from more than ten years back, which renders our conclusions only valid for that particular period. One should however be careful not to interpret them as representative for current subtitling practices. We partially solved these limitations by compiling a new parallel specialized corpus, but its limited size once again prevented us from drawing watertight conclusions about subtitlers' linguistic choices. Furthermore, the results varied depending on the corpus that was used in the case studies. The compilation of a new corpus with recent subtitle material therefore seems warranted. Not only would it allow future researchers to substantiate or contradict the results of this study, it would also enable them to set up a diachronic study that could investigate whether the amount of CBD in subtitles has increased in a period of ten to fifteen years, which could substantiate hypotheses about destandardization or demotization.

Finally, the qualitative part of this dissertation repeatedly allowed us to explain the quantitative results of the case studies in terms of cognitive entrenchment and salience. This qualitative analysis was, however, rather modest in its design, with only two subtitlers and the head of the subtitling department who were involved in this research step. Furthermore, we did not investigate systematically which language features were perceived as more salient than other. The results of this study should therefore be interpreted as tentative, and additional perception and process-based research is needed to support the observations that were made in this study.

Although the present research encountered a number of limitations, the results from the corpus-based multivariate analyses and the qualitative approach showed that subtitles are the result of attempts to find a delicate linguistic balance between the official language policy, that aims to guarantee the general intelligibility of the TV programs, and

the needs and requirements of the television viewers who often have opposite expectations. Furthermore, subtitlers have to take into consideration the authenticity of the program genre and technical criteria like the limited number of characters and pressure of time. These aspects make television subtitles a very specific translation type and subtitlers' norm-related linguistic choices cannot be dissociated from the interaction with the other elements of the audiovisual product.

Bibliography

Absillis, K., Jaspers, J., & Van Hoof, S. (2012). *De manke usurpator: over Verkavelingsvlaams.* Gent: Academia Press.

Árnason, K. (2003). Icelandic. In W. Vandenbussche & A. Deumert (Eds.), *Germanic Standardizations. Past to Present.* Amsterdam/Philadelphia: John Benjamins, 245-279.

Auer, P. (2005). Europe's sociolinguistic unity, or: A typology of European dialect/standard constellations. *Perspectives on variation: Sociolinguistic, historical, comparative,* 7-42.

Auer, P. (2011). Dialect vs. standard: A typology of scenarios in Europe. In B. Kortmann & J. Van der Auwera (Eds.), *The Languages and Linguistics of Europe: A Comprehensive Guide.* Berlin: De Gruyter, 485-500.

Auer, P., & Spiekermann, H. (2011). Demotisation of the standard variety or destandardisation? The changing status of German in Late Modernity (with special reference to southwestern Germany). In T. Kristiansen & N. Coupland (Eds.), *Standard languages and language standards in a changing Europe.* Oslo: Novus Press, 161- 176.

Baker, M. (2004). A corpus-based view of similarity and difference in translation. *International Journal of Corpus Linguistics, 9*(2), 167-193.

Baker, P. (2006). *Glossary of corpus linguistics.* Edinburgh: University Press.

Baños, R. (2013). 'That is so cool': investigating the translation of adverbial intensifiers in English-Spanish dubbing through a parallel corpus of sitcoms. *Perspectives, 21*(4), 526-542.

Bardovi-Harlig, K. (1987). Markedness and salience in second-language acquisition. *Language Learning 37,* 385-407

Bartsch, R. (1984a). Normentheorie en een sociolinguistische toepassing. *TTT. Interdisciplinair Tijdschrift voor Taal en Tekstwetenschap, 4*(3), 177-196.

Bartsch, R. (1984b). Norms, tolerance, lexical change, and context-dependence of meaning. *Journal of pragmatics, 8*(3), 367-393.

Bartsch, R. (1987). *Norms of language: theoretical and practical aspects* (Vol. 33). London/New York: Longman.

Bauman, R., & Briggs, C. L. (2003). *Voices of modernity: Language ideologies and the politics of inequality* (No. 21). Cambridge University Press.

Beheydt, L. (1991a). Dertig jaar taalprogramma's op de radio. In L. Beheydt (Ed.), *Taal en omroep.* 's-Gravenhage: Stichting Bibliographia Neerlandica, 37-52.

Beheydt, L. (1991b). Dertig jaar taalprogramma's op de televisie. In L. Beheydt (Ed.), *Taal en omroep.* 's-Gravenhage: Stichting Bibliographia Neerlandica, 53-63.

Beheydt, L. (1993). De toekomst van het Nederlands. Een nieuwe taalstrijd? In Beheydt et al. (Eds.), *Tussen taal en staat: vooruitkijken voor Vlaanderen.* Leuven: Davidsfonds, 75-94.

Bell, A. (2011). Leaving Home: De-europeanisation in a post-colonial variety of broadcast news language. *Standard Languages and Language Standards in a Changing Europe.* Oslo: Novus Press, 177-198.

Bernardini, S., & Ferraresi, A. (2011). Practice, Description and Theory Come Together: Normalization or Interference in Italian Technical Translation? *Meta, 56*(2), 226-246.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Pearson Education.

Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.

Bishop, H., Coupland, N., & Garrett, P. (2005). Conceptual accent evaluation: Thirty years of accent prejudice in the UK. *Acta Linguistica Hafniensia 37*, 131–154.

Brannan, M. J., & Oultram, T. (2012). Participant Observation 17. In G. Symon & C. Cassell (Eds.), *Qualitative organizational research: Core methods and current challenges*. London: Sage, 296-313.

Bruls, E., & Kerkman, E. (1989). Beeld, Spraak en Schrift: Ondertiteling van Films en Televisieprogramma's in Nederland. In K. Dibbets et al. (Eds.), *Jaarboek mediageschiedenis*. Amsterdam: Stichting Mediageschiedenis 1, 165–182.

Bybee, J. (2003). *Phonology and language use* (Vol. 94). Cambridge: University Press.

Cavalheiro, L. (2008). Linguistic variation in subtitling for audiovisual media in Portugal. *Linguistica Antverpiensia 7*, 17–28.

Chiaro, D. (2009). Issues in Audiovisual Translation. In J. Munday (Ed.), *The Routledge Companion to Translation Studies*. London: Routledge, 141–165.

Clyne, M. (1992). *Pluricentric languages. Different norms in different nations*. Berlin/New York: Mouton de Gruyter.

Cohen, J. (1968) Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* 70, 213-220.

Coupland, N., & Bishop, H. (2007). Ideologised values for British accents. *Journal of Sociolinguistics 11*, 74–93.

Coupland, N. (2010). Language, ideology, media and social change. In K. Junod & D. Maillat (Eds.), *Performing the self*. Tübingen: Gunter Narr, 127-151.

Coupland, N., & Kristiansen, T. (2011). SLICE: Critical perspectives on language (de)standardisation. *Standard languages and language standards in a changing Europe*, 11-35.

Coupland, N. (2014). Sociolinguistic change, vernacularization and broadcast British media. In J. Androutsopoulos (Ed.), *Mediatisation and sociolinguistic change*. Berlijn/Boston: Walter de Gruyter, 67-96.

Coupland, N. (2017). Styling syncretic bilingualism on Welsh-language TV: Madamrygbi. In J. Mortensen, N. Coupland, & J. Thogersen (Eds.), *Style, Mediation, and Change: Sociolinguistic Perspectives on Talking Media*. Oxford: Oxford University Press, 77-99.

Creeber, G. (2008). *The television genre book* (2nd ed.). London: Palgrave Macmillan

Čulo, O., Hansen-Schirra, S., & Nitzke, J. (2017). Contrasting terminological variation in post-editing and human translation of texts from the technical and medical domain. In G. De Sutter, M.-A. Lefer, & I. Delaere (Eds.), *Empirical Translation Studies: New Methodological and Theoretical Traditions*, 183-206.

Dayrell, C. (2007). A quantitative approach to compare collocational patterns in translated and non-translated texts. *International Journal of Corpus Linguistics, 12*(3), 375-414.

De Caluwe, J. (2002). Tien stellingen over functie en status van tussentaal in Vlaanderen. In J. De Caluwe, D. Geeraerts, S. Kroon, V. Mamadouh, R. Soetaert, L. Top & T. Vallen (Eds.), *Taalvariatie and Taalbeleid. Bijdragen aan het taalbeleid in Nederland en Vlaanderen*. Antwerpen/Apeldoorn: Garant, 57–67.

De Caluwe, J. (2005). Conflicting language conceptions within the Dutch speaking part of Belgium. In *Standardvariationen und Sprachideologien in verschiedenen Sprachkulturen der Welt/Standard Variations and Language Ideologies in Different Language Cultures around the World*. Bern: Peter Lang, 53-58.

De Caluwe, J. (2006). Tussentaal als natuurlijke omgangstaal in Vlaanderen. *Studia Germanica Gandensia,* 19-34.

De Caluwe, J. (2009). Tussentaal wordt omgangstaal in Vlaanderen. *Nederlandse taalkunde, 14*(1), 8-25.

De Caluwe, J., & Van Renterghem, E. (2011). Regiolectisering en de opkomst van tussentaal in Vlaanderen. *Taal en Tongval, 63*(1), 61-77.

De Caluwe, J. (2012a). Dutch in Belgium: Facing multilingualism in a context of regional monolingualism and Standard Language Ideology. In M. Hüning, U. Vogl, O. Moliner, F. Grin & G. Lüdi (Eds.), *Standard languages and multilingualism in European history.* Amsterdam/Philadelphia: John Benjamins, 259-282

De Caluwe, J. (2012b). Dutch as a bi-centric language: a lexicographic (r)evolution. In R. Muhr (Ed.), *Non-dominant varieties of pluricentric languages: getting the picture.* Frankfurt am Main: Peter Lang Verlag, 141-152.

De Decker, B. (2014). *De chattaal van Vlaamse tieners: een taalgeografische analyse van Vlaamse (sub) standaardiseringsprocessen tegen de achtergrond van de internationale chatcultuur.* Doctoral dissertation, Universiteit Antwerpen.

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

De Ridder, F. (2007). *Vlaamse tussentaal op televisie: een analyse van het fictieve reportageprogramma 'In De Gloria'.* Unpublished Masters' Dissertation, Ghent University.

De Ridder, R. (2015). *'Arme rijke taal'. Audiovisual translation and minority language planning in pluricentric language areas: a case study of Flemish public service broadcasting.* Doctoral dissertation, Dublin City University.

De Sutter, G., Delaere, I., & Plevoets, K. (2012a). Lexical lectometry in corpus-based translation studies. Combining profile-based correspondence analysis and logistic regression modeling. In M. Oakes & M. Ji (Eds.), *Quantitative Methods in Corpus-based Translation Studies. A practical guide to descriptive translation research.* Amsterdam/Philadelphia: John Benjamins Publishing Company, 325-345.

De Sutter, G., Goethals, P., Leuschner, T., & Vandepitte, S. (2012b). Towards methodologically more rigorous corpus-based translation studies. *Across Languages and Cultures, 13*(2), 137-143.

De Sutter, G., Lefer, M.-A., & Delaere, I. (2017). *Empirical Translation Studies: New Methodological and Theoretical Traditions.* Berlin: Mouton De Gruyter.

Den Boon, T., & Geeraerts, D. (2005). *Van Dale: Groot Woordenboek der Nederlandse Taal*: 3 Dl. (14ᵉ editie). Utrecht: Van Dale Lexicografie.

Delaere, I., De Sutter, G., & Plevoets, K. (2012). Is translated language more standardized than non-translated language?: Using profile-based correspondence analysis for measuring linguistic distances between language varieties. *Target. International Journal of Translation Studies, 24*(2), 203-224.

Delaere, I., & De Sutter, G. (2013). Applying a multidimensional, register-sensitive approach to visualize normalization in translated and non-translated Dutch. *Belgian Journal of Linguistics, 27*(1), 43-60.

Delaere, I. (2015). *Do translations walk the line?: visually exploring translated and non-translated texts in search of norm conformity.* Doctoral dissertation, Ghent University.

Delarue, S. (2016). *Bridging the policy-practice gap: how Flemish teachers' standard language perceptions navigate between monovarietal policy and multivarietal practice.* Doctoral dissertation, Ghent University.

Deumert, A., & Vandenbussche, W. (2003a). *Germanic standardizations: past to present* (Vol. 18). Amsterdam/Philadelphia: John Benjamins Publishing.

Deumert, A., & Vandenbussche, W. (2003b). Research directions in the study of language standardization. In A. Deumert & W. Vandenbussche (Eds.), *Germanic standardizations: Past to present.* Amsterdam/Philadelphia: John Benjamins Publishing, 455-470.

Dewulf, B., & Saerens, G. (2000). *Stijlboek Teletekstondertiteling*, internal document VRT-teletekstredactie.

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

Díaz-Cintas, J. (2013). Theory, practice and research. In C. Millán & F. Bartrina (Eds), *The routledge handbook of translation studies*, London/New York: Routledge.

Díaz-Cintas & Remael, A. (2007). *Audiovisual translation: subtitling.* London/New York: Routledge.

Diwersy, S., Evert, S., & Neumann, S. (2014). A weakly supervised multivariate approach to the study of language variation. In B. Szmrecsanyi & B. Waelchli (Eds.), *Aggregating Dialectology, Typology, and Register Analysis. Linguistic Variation in Text and Speech.* Berlin: De Gruyter.

Doens, E. (2000). *Kwalitatief Onderzoek van Ondertiteling voor Doven en Slechthorenden.* VRT studiedienst.

Duranti, A. (1997). *Linguistic Anthropology.* Cambridge: University Press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis (revised edition). Overview of Methodology of Protocol Analysis.* Cambridge: MIT Press.

Everaert, S. (1998). *De opmars van tussentaal. Een onderzoek naar registervariatie bij de openbare omroep.* Unpublished Masters' Dissertation, Ghent University.

Evert, S., & Neumann, S. (2017). The impact of translation direction on characteristics of translated texts: A multivariate analysis for English and German. In G. De Sutter, M.-A. Lefer, & I. Delaere (Eds.), *Empirical Translation Studies: New Methodological and Theoretical Traditions*, 47-80.

Ferguson, C. A. (1959). Diglossia. *Word, 15*(2), 325-340.

Fishman, J. A. (1967). Bilingualism with and without diglossia; diglossia with and without bilingualism. *Journal of social issues, 23*(2), 29-38.

Gal, S. & Irvine, J. T. (1995). The boundaries of languages and disciplines: How ideologies construct difference. *Social research*, 967-1001.

Gambier, Y. (2012). The position of audiovisual translation studies. In C. Millán & F. Bartrina, (Eds.) *The Routledge Handbook of Translation Studies.* London: Routledge, 45-59.

Garrett, P., Selleck, C., & Coupland, N. (2011). English in England and Wales: multiple ideologies. In T. Kristiansen & N. Coupland (Eds.), *Standard languages and language standards in a changing Europe.* Oslo: Novus Press. 57-65.

Geeraerts, D., Grondelaers, S., & Speelman, D. (1999). *Convergentie en divergentie in de Nederlandse woordenschat: een onderzoek naar kleding-en voetbaltermen.* PJ Meertens-Instituut.

Geeraerts, D. (1999). De Vlaamse taalkloof. *Over Taal 38*, 30–34.

Geeraerts, D., Penne, A., & Vanswegenoven, V. (2000). *Thuis*-taal en *Familie*-taal: taalgebruik in Vlaamse soaps. In S. Gillis, J. Nuyts & J.Taeldeman (Eds.). *Met taal om de tuin geleid. Opstellen voor Georges De Schutter.* Antwerpen: University of Antwerp, 161-170.

Geeraerts, D. (2001). Een zondagspak? Het Nederlands in Vlaanderen: gedrag, beleid, attitudes. *Ons Erfdeel 44*, 337–344.

Geeraerts, D. & Van de Velde, H. (2013). Supra-regional characteristics of colloquial Dutch. In F. Hinskens & J. Taeldeman (Eds.), *Language and space: Dutch.* Berlin: De Gruyter/Mouton, 532-555.

Geeraerts, D., & Speelman, D. (2014). A lectometric definition of demotisation and destandardisation. Retrieved from <https://lirias.kuleuven.be/bitstream/123456789/462723/2/Geeraerts_2014pres_destandardization_demotization.pdf>

Geerts, G., & den Boon, T. (1999). *Van Dale Groot woordenboek der Nederlandse Taal: 3 Dl.* (13ᵉ editie). Utrecht: Van Dale Lexicografie.

Ghyselen, A.-S. (2009). Ne zelfzekere leraar of gewoon nen enthousiaste mens? Een matched-guise onderzoek naar de attitude tegenover tussentaal bij West-Vlamingen. *Taal en Tongval, 61*(2), 83-113.

Ghyselen, A.-S. (2016). *Verticale structuur en dynamiek van het gesproken Nederlands in Vlaanderen: een empirische studie in Ieper, Gent en Antwerpen.* Doctoral dissertation, Ghent University.

Ghyselen, A.-S., Delarue, S., & Lybaert, C. (2016). Studying standard language dynamics in Europe. *Taal en Tongval, 68*(2), 75-91.

Ghyselen, A.-S., & Van Keymeulen, J. (2016). Implicational scales in colloquial Belgian Dutch. *Dialectologia et Geolinguistica, 24*(1), 62-82.

Giddens, A. (1984) T*he Constitution of Society. Outline of the Theory of Structuration.* Cambridge: Polity Press.

Gielen, I., & d'Ydewalle, G. (1989). Hoe worden ondertitelde televisieprogramma's bekeken? *Psycholoog, 24*(9), 425-431.

Giles, H. (1970). Evaluative reactions to accents. *Educational Review* 22, 211–227.

Gloy, K. (1975). *Sprachnormen I. Linguistische und soziologische Analysen.* Stuttgart/Bad Cannstatt: Frommann/Holzboog.

Goossens, J. (2000). De toekomst van het Nederlands in Vlaanderen. *Ons erfdeel, 43*(1), 2-13.

Greenacre, M. (2007). *Correspondence analysis in practice* (2nd edition). Boca Raton: Chapman & Hall/CRC.

Grice, H. P. (1969). Utterer's meaning and intention. *The philosophical review, 78*(2), 147-177.

Gries, S. T. (2010). Methodological skills in corpus linguistics: a polemic and some pointers towards quantitative methods. In T. Harris & M. Moreno Jaén (Eds.), *Corpus linguistics in language teaching.* Frankfurt a.M.: Peter Lang, 121-146.

Grondelaers, S., & Van Hout, R. (2011). The standard language situation in the Low Countries: Top-down and bottom-up variations on a diaglossic theme. *Journal of Germanic Linguistics, 23*(3), 199-243.

Grondelaers, S., Van Hout, R., & Speelman, D. (2011). A perceptual typology of standard language situations in the Low Countries. *Standard languages and language standards in a changing Europe*, 199-222.

Grondelaers, S., & Kristiansen, T. (2013). On the need to access deep evaluations when searching for the motor of standard language change. In T. Kristiansen & S. Grondelaers (Eds.), *Language (De)standardisation in Late Modern Europe: Experimental Studies.* Oslo: Novus Press, 9-52.

Grondelaers, S., & Speelman, D. (2013). Can speaker evaluation return private attitudes towards stigmatised varieties? Evidence from emergent standardisation in Belgian Dutch. *Language (De) standardisations in Late Modern Europe: Experimental Studies*, 171-191.

Grondelaers, S., & van Hout, R. (2016). How (in) coherent can standard languages be? A perceptual perspective on co-variation. *Lingua 172*, 62-71.

Grondelaers, S., van Hout, R., & van Gent, P. (2016). Destandardization is not destandardization. *Taal en Tongval, 68*(2), 119-149.

Haarmann, H. (1997). Sprachstandardisierung — Eine kulturanthropologische Konstante. In K. J. Mattheier & E. Radtke (Eds), *Standardisierung und Destandardisierung europäischer Nationalsprachen.* Frankfurt: Lang, 259–290.

Halverson, S. L. (2017). Gravitational pull in translation. Testing a revised model. *Empirical Translation Studies: New Methodological and Theoretical Traditions*, 9-45.

Hamaida, L. (2007). Subtitling slang and dialect. *Proceedings of the MuTra 2007-LSP translation scenarios conference*, 1-11.

Hatim, B., & Munday, J. (2004). *Translation. An advanced resource book.* New York: Routledge.

Hedin, T. (2009). Codification and Language Norm in Czech—Examples from a Corpus Analysis of Spoken Language in Television. *Swedish Contributions to the Fourteenth International Congress of Slavists*, 33-46.

Hendrickx, R. (1998). *Het taalcharter.* VRT-Taalnet. Retrieved from <http://www.vrt.be/taal/taalcharter>.

Hendrickx, R. (2001). Laat ons ne keer te goei naar onszelf luisteren. *VRT. Taalsignaal voor 6 miljoen Vlamingen.* Retrieved from <http://www.vrt.be/taal/laat-ons-ne-keer-te-goei-naar-onszelf-luisteren>.

Hendrickx, R. (2003a). *Stijlboek VRT.* Utrecht: Van Dale Lexicografie.

Hendrickx, R. (2003b). Wat zegt ie? VRT-Taalnet. Retrieved from <https://taal.vrt.be/wat-zegt-ie?print=1>.

Hendrickx, R. (2007). *Taal bij de VRT.* VRT-Taalnet. Retrieved from <http://taal.site/teksten/taalcharter2007.pdf>.

Hendrickx, R. (2011). *Nederlands in het Nederlands ondertitelen.* VRT-Taalnet. Retrieved from <http://taal.vrt.be/nederlands-in-het-nederlands-ondertitelen>.

Hendrick, R. (2012). *Taalcharter 2012.* VRT-Taalnet. Retrieved from <https://taal.vrt.be/taalcharter-2012>.

Hilton, N. H. (2010). *Regional dialect levelling and language standards: changes in the Hønefoss dialect.* Doctoral dissertation, University of York.

Hodson, J. (2014). *Dialect in film and literature.* New York: Palgrave Macmillan.

Impe, L. (2006). *Een attitudineel mixed-guise onderzoek naar tussentaal in Vlaanderen.* Unpublished Masters' Dissertation, University of Leuven.

Impe, L., & Speelman, D. (2007). Vlamingen en hun (tussen) taal. Een attitudineel mixed guiseonderzoek. *Handelingen van de Koninklijke Maatschappij voor Taal-en Letterkunde en Geschiedenis 16*, 109-128.

Jacobs, G., Van Hout, T., & Van Praet, E. (2011). *Discursive perspectives on news production. Journal of Pragmatics.* Amsterdam: Elsevier.

Jahr, E. H. (2003). Norwegian. In A. Deumert & W. Vandenbussche (Eds.), *Germanic standardizations. Past to Present.* Amsterdam/Philadelphia: John Benjamins, 331-353.

Janssens, G., & Marynissen A. (2008). *Het Nederlands vroeger en nu.* (3rd ed.) Leuven/Voorburg: Acco.

Jaspers, J. (2001). Het Vlaamse stigma. Over tussentaal en normativiteit. *Taal en Tongval, 53*(2), 129-153.

Jaspers, J., & Brisard, F. (2006). Verklaringen van substandaardisering: tussentaal als gesitueerd taalgebruik. *Leuvense bijdragen 95*, 35-70.

Jaspers, J., & Meeuwis, M. (2006). Playing with language in a media-promoted standardisation context: public broadcasting and adolescent linguistic practices in Flanders. *Acta Academica* (Supplement 2), 111-140.

Jaspers, J. (2010). Style and styling. *Sociolinguistics and language education*, 177-204.

Jaspers, J., & Van Hoof, S. (2015). Ceci n'est pas une tussentaal: Evoking standard and vernacular language through mixed Dutch in Flemish telecinematic discourse. *Journal of Germanic Linguistics, 27*(1), 1-44.

Jenset, G. B., & McGillivray, B. (2012). Multivariate analyses of affix productivity in translated English. *Quantitative methods in corpus-based translation studies*, 301-323.

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

Kennedy, G. D. (1998). *An introduction to corpus linguistics.* London: Longmann.

Kilborn, R. (1993). Speak my Language: Current Attitudes to Television Subtitling and Dubbing. *Media, Culture and Society 15*, 641–660.

Koolstra, C. M., Peeters, A. L., & Spinhof, H. (2002). The pros and cons of dubbing and subtitling. *European Journal of Communication, 17*(3), 325-354.

Kristiansen, T. (2001). Two Standards: One for the media and one for the school. *Language Awareness 10*(1), 9–24.

Kristiansen, T. (2003a). Language attitudes and language politics in Denmark. *International Journal of the Sociology of Language*, 57-72.

Kristiansen, T. (2003b). The youth and the gatekeepers: reproduction and change in language norm and variation. In J. K. Androutsopoulos & A. Georgakopoulou (Eds.) *Discourse Constructions of Youth Identities.* Amsterdam: Benjamins, 279–302.

Kristiansen, T. (2009). The macro-level social meanings of late-modern Danish accents. *Acta linguistica hafniensia, 41*(1), 167-192.

Kroskrity, P. V. (2000). *Regimes of language: Ideologies, polities, and identities.* James Currey Publishers.

Kruger, H., & van Rooy, B. (2012). Register and the features of translated language. *Across Languages and Cultures, 13*(1), 33-65.

Kruger, H. (2017). The effects of editorial intervention. Implications for studies of the features of translated language. In G. De Sutter, M.-A. Lefer, & I. Delaere (Eds.), *Empirical Translation Studies: New Methodological and Theoretical Traditions*, 113-156.

Kruger, H. (in press). *That* again: A multivariate analysis of the factors conditioning syntactic explicitness in translated English. *Across Languages and Cultures.*

KU Leuven & VRT (1999). *Project taalhantering VRT 1998-1999. Eindevaluatie.* VRT-archief.

Labov, W. (1966). *The social stratification of English in New York city.* Washington DC: Center for Applied Linguistics.

Langacker, R. W. (1987). F*oundations of Cognitive Grammar. Theoretical prerequisites* (Vol. 1). Stanford: Stanford University Press.

Laviosa, S. (1998). Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta 43*, 557-570.

Lebbe, D. (1996). *Schoon Vlaams. Een onderzoek naar syntactische en morfo-syntactische aspecten van 'tussentaal' zoals die door Ieperlingen wordt gesproken.* Unpublished Masters' Dissertation, Ghent University.

Lefevere, E. (2011). *Tussentaal in de Vlaamse media. Een onderzoek naar het taalgebruik van televisiepresentatoren.* Unpublished Masters' Dissertation, Ghent University.

Lemahieu, G. (2008). *Tussentaal in spontane conversaties van Vlaamse jongvolwassenen (18- tot 24-jarigen).* Unpublished Masters' Dissertation, Ghent University.

Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States.* London/New York: Routledge.

Lüdi, G. (2012). Traces of monolingual and plurilingual ideologies in the history of language policies in France. In M. Hüning, U. Vogl, O. Moliner, F. Grin & G. Lüdi (Eds.), *Standard languages and multilingualism in European history.* Amsterdam/Philadelphia: John Benjamins, 205-230.

Luhmann, N. (1972). *Rechtssoziologie*, 2 Bde. Reinbek bei Hamburg: Rowohlt.

Luyken, G., Herbst, T., Langham-Brown, J., Reid, H., & Spinhof, H. (1991). *Overcoming Language Barriers in Television: Dubbing and Subtitling for the European Audience.* Manchester: European Institute for the Media.

Lybaert, C. (2014a). *Het gesproken Nederlands in Vlaanderen: percepties en attitudes van een spraakmakende generatie.* Doctoral dissertation, Ghent University.

Lybaert, C. (2014b). Perceptie van tussentaal in het gesproken Nederlands in Vlaanderen. *Nederlandse Taalkunde*, *19*(2), 185–220.

Lybaert, C. (2015). De ideologische positie van Algemeen Nederlands in Vlaanderen. *Taal en Tongval 67*(1), 97-134.

Lybaert, (2017). A direct discourse-based approach to the study of language attitudes: The case of tussentaal in Flanders. *Language Sciences 59*, 93-116.

Macken, L., De Clercq, O. & Paulussen, H. (2011). Dutch Parallel Corpus: A Balanced, Copyrightcleared Parallel Corpus. *Meta 56*(2), 374–390.

Malmkjaer, K. (1997). Punctuation in Hans Christian Andersen's stories and in their translations into English. In F. Poyatos (Ed.), *Nonverbal Communication and Translation.* Philadelphia: John Benjamins.

Marklund Sharapova, E. (2000). *Implicit and explicit norm in contemporary Russian verbal stress.* Doctoral dissertation, Acta Universitatis Upsaliensis.

Martin, W. (2005). *Het Belgisch-Nederlands anders bekeken: het Referentiebestand Belgisch-Nederlands (RBBN).* Tech. rept. Vrije Universiteit Amsterdam, Amsterdam, the Netherlands.

Mattheier, K. J. (1997). Über destandardiseiering, Umstandardisierung and Standardisierung in modernen Europäischen Standardsprachen. In K. Mattheier & E. Radtke (Eds.) *Standardisierung und Destandardisierung Europäischer Nationalsprachen.* Frankfurt: Peter Lang Verlag, 1-9.

Mattheier, K. J. (2003). German. In A. Deumert & W. Vandenbussche (Eds.), *Germanic standardizations. Past to Present.* Amsterdam/Philadelphia: John Benjamins, 211-244.

Mauranen, A., & Kujamäki, P. (2004). *Translation universals: Do they exist?* (Vol. 48). Amsterdam/Philadelphia: John Benjamins Publishing.

McIlvenny, P., Mettovaara, S., & Tapio, R. (1992). I really wanna make you laugh: Stand-up comedy and audience response. *Folia, fennistica and linguistica: Proceedings of the annual Finnish linguistic symposium May*, 225-247.

Mera, M. (1999). Read my lips: Re-evaluating subtitling and dubbing in Europe. *Links & Letters 6*, 73-85.

Milroy, J., & Milroy, L. (1985). *Authority in language: Investigating language prescription and standardisation.* London: Routledge.

Milroy, J. (2001). Language ideologies and the consequences of standardization. *Journal of sociolinguistics, 5*(4), 530-555.

Mugglestone, L. (2003). *Talking proper. The rise and fall of the English accent as a social symbol.* Oxford: Oxford University Press.

Myers, F. R. (1986). *Pintupi Country Pintupi Self: Sentiment, Place, and Politics among Western Desert Aborigines.* Washington: Smithsonian Institution Press.

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

Nikken, P., & Friebel, C. (1990). *Tussen Kijkcijfers en Boodschap. Omroepmedewerkers over Jeugdprogrammering.* Leiden: Rijksuniversiteit Leiden.

Nowak, S., & Rüger, S. (2010). How reliable are annotations via crowdsourcing: a study about inter-annotator agreement for multi-label image annotation. *Proceedings of the international conference on Multimedia information retrieval*, 557-566.

Oakes, M., & Ji, M. (2012). *Quantitative Methods in Corpus-Based Translation Studies* (Vol. 51). Amsterdam: John Benjamins.

Odendaal, G. (2013). Restandardisation defined as democratising language planning. *Stellenbosch Papers in Linguistics Plus 42*, 183-203.

Olohan, M., & Baker, M. (2000). Reporting that in Translated English: Evidence for Subconscious Processes of Explicitation? *Across Languages and Cultures, 1*(2), 141-158.

Oostdijk, N., Reynaert, M., Hoste, V., & Schuurman, I. (2013). The construction of a 500-million-word reference corpus of contemporary written Dutch. In *Essential speech and language technology for Dutch.* Berlin/Heidelberg: Springer, 219-247

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

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

Plevoets, K. (2013). De status van de Vlaamse tussentaal: een analyse van enkele socio-economische determinanten. *Tijdschrift voor Nederlandse Taal-en Letterkunde, 129*(3), 191-233.

Plevoets, K. (2015). Corregp: Functions and Methods for Correspondence Regression. Ghent University. Retrieved from
<https://cran.rproject.org/web/packages/corregp/index.html>.

Prieels, L. (2013). Tussentaal in tv-programma's. In J. De Caluwe, S. Delarue, A.-S. Ghyselen & C. Lybaert (Eds.) *Tussentaal: over de talige ruimte tussen dialect en standaardtaal in Vlaanderen*, 35-49.

Pym, A. (2005). Explaining explicitation. *New trends in translation studies. In honour of Kinga Klaudy*, 29-34.

Pym, A. (2008). On Toury's laws of how translators translate. In A. Pym, M. Shlesinger & D. Simeoni (Eds.), *Beyond Descriptive Translation Studies. Investigations in Honor of Gideon Toury*. Amsterdam/Philadelphia: Benjamins, 311-328.

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

Ramos Pinto, S. (2009). How important is the way you say it?: A discussion on the translation of linguistic varieties. *Target. International Journal of Translation Studies, 21*(2), 289-307.

Ramos Pinto, S. (2017). Film, dialects and subtitles: an analytical framework for the study of non-standard varieties in subtitling. *The Translator*, 1-18.

Reiczigel, J. (1996). Bootstrap tests in correspondence analysis. *Applied Stochastic Models in Business and Industry, 12*(2), 107-117.

Remael, A. (2003). Mainstream narrative film dialogue and subtitling. *The Translator, 9*(2), 225–247.

Remael, A. (2007). Vertaling in beeld: aspecten van audiovisuele vertaling en ondertiteling. *Methodisch kijken: aspecten van onderzoek naar film-en beeldcultuur*, 183-205.

Remael, A., De Houwer, A., & Vandekerckhove, R. (2008). Intralingual open subtitling in Flanders: Audiovisual translation, linguistic variation and audience needs. *Journal of Specialized Translation, 10*, 76–105.

Reynaert, M., Oostdijk, N., De Clercq, O., van den Heuvel, H., & de Jong, F. (2010). Balancing SoNaR: IPR versus processing issues in a 500-million-word written Dutch reference corpus. *Seventh conference on international language resources and evaluation,* LREC '10, 2693–2698, 19–21 May 2010, Malta.

Rosa, A. A. (1994). The Centre and the Edges. Linguistic Variation and Subtitling Pygmalion into Portuguese. *Translation and the (Re) location of Meaning,* 317-337.

Rosa, A. A. (2001). Features of oral and written communication in subtitling. In Y. Gambier, & H. Gottlieb (Eds.), *(Multi)media translation: Concepts, practices and research*. Amsterdam/Philadelphia: Benjamins Translation Library, 213–222.

Rosa, A. A. (2016). Descriptive translation studies of audiovisual translation. *Target. International Journal of Translation Studies, 28*(2), 192-205.

Rosewarne, D. (1984). Estuary English. Times Educational Supplement 19. Retrieved from <http://www.phon.ucl.ac.uk/home/estuary/rosew.htm>.

Rutten, G., & Vosters, R. (2010). Spellingsnormen in het Zuiden. Standaardisatie van het geschreven Nederlands in de achttiende en negentiende eeuw. In M. van der Wal & E. Francken (Eds.), *Standaardtalen in beweging*. Amsterdam: Stichting Neerlandistiek VU, 27-48.

Rutter, J. (1997). Stand-up as interaction: Performance and audience in comedy venues. Unpublished dissertation, University of Salford.

Rybicki, J. (2012). The great mystery of the (almost) invisible translator. In J. Meng & M. Oakes (Eds.), *Quantitative Methods in Corpus-Based Translation Studies*, Amsterdam/Philadelphia: John Benjamins, 231-248.

Rys, K., & Taeldeman, J. (2007). Fonologische ingrediënten van Vlaamse tussentaal. *Tussen taal, spelling en onderwijs. Essays bij het emeritaat van Frans Daems*, 1-8.

Saman, V. (2003). *Studie van het taalgebruik in reclamespots op de radio in Vlaanderen van 1991 tot 2001. Tussentaal in opmars.* Unpublished Masters' Dissertation, Universiteit Gent.

Sandøy, H. (2011). Language culture in Norway: a tradition of questioning standard language norms. In T. Kristiansen & N. Coupland (Eds.), *Standard languages and language standards in a changing Europe*. Oslo: Novus Press, 119-126.

Schiffer, S. R. (1972). *Meaning.* Oxford: Clarendon Press.

Schmid, H.-J. (2007). Entrenchment, salience, and basic levels. In: D. Geeraerts & H. Cuyckens (Eds.), *The Oxford handbook of cognitive linguistics.* Oxford: University Press.

Schmid, H.-J. (2010). Does frequency in text instantiate entrenchment in the cognitive system. *Quantitative methods in cognitive semantics: Corpus-driven approaches*, 101-133.

Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching, 10*(1-4), 209-232.

Silverstein, M. (1979). Language structure and linguistic ideology. In R. Clyne, W. Hanks & C. Hofbauer (Eds.), *The Elements: A parasession on linguistic units and levels.* Chicago: Chicago Linguistic Society, 193-247.

Slembrouck, S., & Van Herreweghe, M. (2004). Teletekstondertiteling en tussentaal: de pragmatiek van het alledaagse. In J. De Caluwe, M. Devos, J. Van Keymeulen & G. De Schutter (Eds.), *Taeldeman, man van de taal, schatbewaarder van de taal: liber amicorum Johan Taeldeman.* Ghent: Academia press, 853-876.

Sleurs, K., & Jacobs, G. (2005). Beyond preformulation: an ethnographic perspective on press releases. *Journal of Pragmatics, 37*(8), 1251-1273.

Smakman, D. (2006). *Standard Dutch in the Netherlands: A sociolinguistic and phonetic description.* Utrecht: Lot.

Smith, J. (1996). *An Historical Study of English. Function, form and change*. London: Routledge.

Soukup, B. (2016). Doing 'speaking the (non-) standard'in the media. *Taal en Tongval, 68*(2), 151-172.

Speelman, D., Grondelaers, S., & Geeraerts, D. (2003). Profile-based linguistic uniformity as a generic method for comparing language varieties. *Computers and the Humanities, 37*(3), 317-337.

Spinhof, H., & Peeters, A.L. (1999). *Opinies over Nasynchroniseren en Ondertitelen. Report No. P99-160.* Hilversum: NOS.

Swann, J., Deumert, A., Lillis, T., & Mestrie, R. (2004). *A Dictionary of Sociolinguistics.* Edinburgh: Edinburgh University Press.

Szarkowska, A. (2005). The power of film translation. *Translation journal, 9*(2), 2005.

Szarkowska, A. (2016). *Report on the results of an online survey on subtitle presentation times and line breaks in interlingual subtitling. Part 1: Subtitlers.* London.

Taeldeman, J. (1992). Welk Nederlands voor Vlamingen?. *Nederlands van nu, 40*, 33-51.

Taeldeman, J. (2008). Zich stabiliserende grammaticale kenmerken in Vlaamse tussentaal. *Taal en Tongval*, *60*, 26-50.

Thelander, M. (2011). Standardisation and standard language in Sweden. *Standard languages and language standards in a changing Europe*, 127-133.

Torres Cacoullos, R. & Walker, J. A. (2009). On the Persistence of Grammar in Discourse Formulas: A Variationist Study of that. *Linguistics 47* (1), 1–43.

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

Trudgill, P., & Giles, H. (1978). Sociolinguistics and linguistic value judgments: Correctness, adequacy and aesthetics. In F. Coppieters & D.L. Goyvaerts (Eds.) *Functional Studies in Language and Literature.* Gent: Story-Scientia, 167–190.

Ullmann-Margalit, E. (2015). *The emergence of norms* (Vol. 11). Oxford: University Press.

Valentini, C. (2013). Phrasal verbs in Italian dubbed dialogues: a multimedia corpus-based study. *Perspectives, 21*(4), 543-562.

Van Bezooijen, R. (2002). Aesthetic evaluation of Dutch. Comparisons across dialects, accents, and languages. In D. Long & D. R. Preston (Eds.) *Handbook of Perceptual Dialectology* (Vol. 2). Amsterdam: John Benjamins, 13–30.

Van Bree, C. (2000). Taalbewustzijn, taalverandering en regionale variatie. Een benadering. *Taal en Tongval, 52*(1), 22–46.

van Dale, J.H., den Boon, T., Hendrickx, R., & Van der Sijs, N. (2015). *Van Dale: Groot woordenboek van de Nederlandse taal*: 3Dl (15e editie). Utrecht: Van Dale Uitgevers.

Van de Velde, H. (1996). Verkavelingsvlaams. Wat is mij dat nu. In R. van Hout & J. Kruijsen (Eds.) *Taalvariaties. Toonzettingen en modulaties op een thema.* Dordrecht: Foris Publications, 261-271.

Van Gijsel, S., Speelman, D., & Geeraerts, D. (2008). Style shifting in commercials. *Journal of Pragmatics*, *40*, 205-226.

Van Hoof, S. (2010). Taalvariatie op de Vlaamse openbare omroep: Substandaardisering in fictieprogramma's vroeger en nu. *Handelingen van de Koninklijke Zuid-Nederlandse Maatschappij voor Taal-en Letterkunde en Geschiedenis, 64*, 167-185.

Van Hoof, S., & Jaspers, J. (2012). Hyperstandaardisering. *Tijdschrift voor Nederlandse Taal- en Letterkunde 128*(2). 97-125.

Van Hoof, S. (2013). *Feiten en fictie: een sociolinguïstische analyse van het taalgebruik in fictieseries op de Vlaamse openbare omroep (1977-2012).* Doctoral Disseration, University of Antwerp.

Van Hoof, S. (2015). *Feiten en fictie. Een sociolinguïstische analyse van het taalgebruik in fictiereeksen op de Vlaamse openbare omroep (1977-2012).* Ghent: Academia Press.

Van Istendael, G. (1989). *Het Belgisch labyrint: de schoonheid der wanstaltigheid.* Amsterdam: Arbeiderspers.

Van Poecke, L., & Van den Bulck, H. (1991). Taal en omroep. Een schets van de evolutie in het taalbeleid van de Vlaamse Openbare Omroep. In L. Beheydt (Ed.), *Taal en omroep.* 's-Gravenhage: Stichting Bibliographia Neerlandia, 83-107.

Vancompernolle, H. (2012). *Normgevoeligheid: attitude van Vlaamse jongeren ten aanzien van het Standaardnederlands, de tussentaal en het dialect* (Vol. 5). Ghent: Academia Press.

Vandekerckhove, P., & Ghyselen, A.-S. (2017). De detectie van saillante taalkenmerken. *Nederlandse Taalkunde, 22*(2), 249-279.

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

Vandekerckhove, R. (2007). 'Tussentaal' as a source of change from below in Belgian Dutch. A case study of substandardization processes in the chat language of Flemish teenagers. *Germanic language histories "from below"(1700-2000).* Berlin: De Gruyter, 189-203.

Vandekerckhove, R., De Houwer A., & Remael, A. (2007). Intralinguale ondertiteling op de Vlaamse televisie: een spiegel voor de taalverhoudingen in Vlaanderen? In D. Sandra, R. Rymenans, P. Cuvelier & P. Van Petegem (Eds.), *Tussen taal, spelling en onderwijs. Essays bij het emiraat van Frans Daems.* Gent: Academia Press, 71–83.

Vandekerckhove, R., De Houwer, A., & Remael, A. (2009). Between language policy and linguistic reality. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA), 19*(4), 609-628.

Vandekerckhove, R., & Nobels, J. (2010). Code eclecticism: Linguistic variation and code alternation in the chat language of Flemish teenagers. *Journal of Sociolinguistics, 14*(5), 657-677.

Vandenbussche, W. (2010). Standardisation through the media. The case of Dutch in Flanders. In P. Gilles, J. Scharloth & E. Ziegler (Eds.), *Variatio delectat: empirische Evidenzen und theoretische Passungen sprachlicher Variation. Herausgegeben für Klaus J. Mattheier zum 65. Geburtstag.* Frankfurt am Main: Peter Lang Verlag, 309- 322.

Vosters, R., & Vandenbussche, W. (2008). Wijzer worden over Willem? Taalgebruik in Vlaanderen ten tijde van het Verenigd Koninkrijk der Nederlanden (1814-1830). *Internationale Neerlandistiek 46*(3). 2-22.

Vosters, R., Rutten, G. J., & van der Wal, M. J. (2010). Mythes op de pijnbank. Naar een herwaardering van de taalsituatie in de Nederlanden in de achttiende en negentiende

eeuw. *Verslagen en Mededelingen van de Koninklijke Academie voor Nederlandse Taal-en Letterkunde* 120, 93-112.

Vosters, R. (2013). Dutch, Flemish or Hollandic? Social and ideological aspects of linguistic convergence and divergence during the United Kingdom of the Netherlands (1815-1830). In E. Barát, P. Studer & J. Nekvapil (Eds.), *Ideological conceptualisations of language. Discourses of linguistic diversity*. Frankfurt am Main: Peter Lang Verlag, 35-54.

VRT Vertaling en ondertiteling (2009). *Stijlboek 888: Teletekstondertiteling* [internal document] Brussel: VRT-teletekstredactie.

Willemyns, R. (2003). Dutch. In A. Deumert & W. Vandenbussche (Eds.), *Germanic standardizations: past to present* (Vol. 18). Amsterdam/Philadelphia: John Benjamins Publishing, 93-125.

Willemyns, R., & Daniëls, W. (2003). *Het verhaal van het Vlaams: de geschiedenis van het Nederlands in de Zuidelijke Nederlanden*. Antwerpen: Standaard Uitgeverij.

Willemyns, R.(2013). *Dutch. Biography of a language*. Oxford: University Press.

Woolard, K. A. (1992). Language ideology: Issues and approaches. *Pragmatics, 2*(3), 235-249.

Zabalbeascoa, P. (2008) The Nature of the Audiovisual Text and its Parameters. In J. Díaz Cintas (Ed.), *The Didactics of Audiovisual Translation.* Amsterdam/Philadelphia: John Benjamins, 21–39.

Zenner, E., Geeraerts, D., & Speelman, D. (2009). Expeditie Tussentaal-Leeftijd, identiteit en context in "Expeditie Robinson". *Nederlandse taalkunde, 14*(1), 26-44.

Zubiria, J. B. (2012). *Mapping the dubbing scene: Audiovisual translation in Basque television. New Trends in Translation Studies* (Vol. 2). Bern: Peter Lang.

Summary

In Corpus-based Translation studies, it is traditionally believed that there is a general tendency to normalize or standardize translations, which implies that translators use more neutral expressions, more conventional and less creative language than non-translators. Nevertheless, the specificity of the language that is used in translations does not only depend on their translational status, but also on a number of other factors such as genre, source language and target audience. In this context, the present dissertation investigated norm-related language variation in subtitling, which constitutes a special type of translations, as subtitles are characterized by a colloquial input and are thus situated between oral and written language. More particularly, we investigated whether subtitlers in Flanders prefer Colloquial Belgian Dutch lexemes and constructions rather than Belgian Standard Dutch lexemes and constructions, and whether these linguistic choices are influenced by the contextual parameters *source language* and *program genre.*

The specific language situation in Flanders makes this research particularly interesting. Next to the official standard language (viz. Belgian Standard Dutch), various non-standard varieties (e.g. Colloquial Belgian Dutch) are widely used in Flanders, both in informal and formal contexts, although they are not accepted by the language authorities. The present study explored how Flemish subtitlers deal with the linguistic tension between the norm-adherent language policy of the Flemish public broadcaster VRT on the one hand, which is largerly oriented toward the use of Belgian Standard Dutch, and the particular language situation in Flanders on the other hand. Given that subtitles are the result of the transformation of spoken language, with its typical colloquial features, into written text, it was particularly interesting to investigate which variety is used by the subtitlers, and how these linguistic choices are affected by different contexts.

To achieve these goals, we compiled several lexical and grammatical profiles, i.e. pairs of language variants, with one being the Belgian Standard Dutch variant (e.g. *handtas* 'hand bag') and the other being the Colloquial Belgian Dutch variant (e.g. *sacoche* 'hand bag'). These profiles were extracted from different corpora, containing interlingual and intralingual subtitles on the one hand and translated and non-translated written text

genres on the other hand. The obtained data were manually validated and annotated for several contextual parameters (*source language*, *program genre*, *target audience*, *program purpose*, and *speaker type*), and subjected to profile-based correspondence analysis, a multivariate statistical technique which allows the visualization of the subtitlers' norm-related language choices in the aforementioned contexts. Finally, a qualitative analysis of the obtained results was done to get insight into subtitlers' attitudes toward VRT's language policy and the language reality in Flanders, which is crucial understand their linguistic choices.

In total, three case studies were carried out to investigate norm-adherence in subtitling on Flemish television. Case study 1 compared the linguistic choices made in written translations and non-translations to the linguistic choices made by subtitlers in order to investigate to what extent subtitlers, translators and writers of original Dutch texts conform to the Standard-Dutch norm. In addition, it was verified whether the contextual factors *source language*, *program genre*, and *speaker type* had an influence on the linguistic choices of the subtitlers. The second case study did not only investigate the extent to which Flemish subtitlers use Colloquial Belgian Dutch variants instead of Belgian Standard Dutch variants, but also whether the subtitles contain more colloquial lexemes than colloquial grammatical constructions. In addition, the influence of the contextual factors *source language* and *program genre* on the subtitlers' lexical and grammatical choices was examined. The final case study addressed the question to what extent Flemish subtitlers reproduce the Belgian-Dutch colloquialisms from the spoken source text in the subtitles by comparing the language used in twenty television programs to the corresponding intralingual subtitles. Furthermore, it was examined whether the subtitlers more often opted for lexical colloquialisms than morphological or syntactic colloquialisms, and whether the program genre influenced these linguistic choices. In addition, subtitlers were interviewed to get more contextual information about the practical as well as the political concerns they have to deal with and to find out why they opt for the reproduction or translation of certain Belgian-Dutch colloquialisms.

Although this dissertation suffered from a number of limitations, the empirical findings from the corpus-based multivariate analyses and the qualitative approach have revealed that subtitles are the result of a delicate process to find a linguistic balance between the official language policy, that aims to guarantee the general intelligibility of the TV programs, and the needs and requirements of the television viewers who often have opposite expectations. Furthermore, the authenticity of the program genre and technical criteria like the limited number of characters and pressure of time have an impact on the subtitlers' linguistic choices.

Samenvatting

Binnen het domein van Corpusgebaseerde Vertaalwetenschap wordt algemeen aangenomen dat vertaalde teksten meer genormaliseerd en meer gestandaardiseerd zijn dan niet-vertaalde teksten. Dat houdt in dat vertalers meer neutrale constructies, meer conventionele en minder creatieve taal gebruiken dan niet-vertalers. Toch wordt het specifieke karakter van vertaalde taal niet enkel bepaald door de vertaalstatus, maar ook door andere factoren zoals het tekstgenre, de brontaal en het doelpubliek. In die context bestudeert dit proefschrift normgerelateerde taalvariatie in ondertiteling, een speciaal vertaaltype dat zich door zijn spreektalige input tussen spreek- en schrijftaal bevindt. We zijn onder meer nagegaan of Vlaamse ondertitelaars een voorkeur hebben voor tussentalige woorden en constructies in plaats van standaardtalige woorden en constructies, en of die keuze beïnvloed wordt door de contextuele factoren *brontaal* en *programmagenre.*

De specifieke taalsituatie in Vlaanderen maakt dit onderzoek uitermate interessant. Naast de officiële standaardtaal (Belgisch Standaardnederlands) worden er in Vlaanderen immers allerlei niet-standaardtalige variëteiten (bv. tussentaal) gebruikt, zowel in informele als in formele situaties, ook al worden die variëteiten niet aanvaard door de officiële taalbeleidsinstanties. Deze studie onderzoekt hoe Vlaamse ondertitelaars omgaan met de talige spanning tussen het normgerichte taalbeleid van de Vlaamse openbare omroep VRT enerzijds en de kenmerkende taalsituatie in Vlaanderen anderzijds. Aangezien ondertitels het resultaat zijn van de transformatie van gesproken taalgebruik in geschreven taal, was het uitermate interessant om na te gaan welke variëteit(en) de ondertitelaars hanteren en in welke mate die keuze wordt beïnvloed door verschillende extralinguïstische contexten.

Om die doelstellingen te bereiken hebben we lijsten met lexicale en grammaticale profielen opgesteld. Een profiel is een set van taalvarianten met minstens één standaardtalige variant (bv. *handtas*) en één tussentalige variant (bv. *sacoche*). Die profielen werden uit verschillende corpora met interlinguale en intralinguale ondertitels en vertaalde en niet-vertaalde teksten onttrokken. De verkregen data werden manueel gevalideerd en geannoteerd aan de hand van verscheidene parameters (*brontaal,*

programmagenre, *doelpubliek*, *programmadoel* en *sprekerstype*) en de dataset werd onderworpen aan een profielgebaseerde correspondentieanalyse. Dat is een statistische techniek die de normgerelateerde taalkeuzes van de ondertitelaars in bovengenoemde contexten visualiseert. Tot slot werden de kwantitatieve resultaten kwalitatief geanalyseerd om een beter beeld te krijgen van de houding van de ondertitelaars ten aanzien van het taalbeleid van de openbare omroep en de Vlaamse taalrealiteit, wat van cruciaal belang was om hun taalkeuzes te kunnen begrijpen.

In totaal voerden we drie gevalsstudies uit om het normconformisme in ondertiteling op de Vlaamse televisie te onderzoeken. In de eerste gevalsstudie vergeleken we de taalkeuzes in vertaalde en niet-vertaalde teksten met de taalkeuzes die ondertitelaars maakten om na te gaan in welke mate ondertitelaars, vertalers en schrijvers van originele Nederlandse teksten de Standaardnederlandse norm volgen. Bovendien gingen we na of de *brontaal*, het *programmagenre* en het *sprekerstype* een invloed hadden op de taalkeuzes van de ondertitelaars. In de tweede gevalsstudie werd niet enkel onderzocht in welke mate Vlaamse ondertitelaars Standaardnederlands of tussentaal gebruikten, maar ook of de ondertitels meer tussentalige woorden dan tussentalige grammaticale constructies bevatten. We onderzochten ook nog of de *brontaal* en het *programmagenre* de lexicale en grammaticale keuzes van de ondertitelaars beïnvloedden. In de laaste gevalsstudie gingen we na in welke mate de ondertitelaars de spreektalige elementen uit de brontekst overnamen door het taalgebruik in twintig televisieprogramma's te vergelijken met de corresponderende intralinguale ondertitels. Daarnaast werd er onderzocht of de ondertitelaars vaker opteerden voor tussentalige lexemen dan voor tussentalige morfologische en syntactische elementen, en of het *programmagenre* een invloed had op die keuzes. Tot slot werden er enkele ondertitelaars geïnterviewd om meer contextuele informatie te krijgen van zowel de praktische als de taalbeleidspolitieke beperkingen waarmee ze te maken krijgen en om te achterhalen waarom ze kiezen voor de overname of de omzetting van bepaalde tussentalige kenmerken.

Hoewel het onderzoek in dit proefschrift werd blootgesteld aan enkele beperkingen, hebben zowel de empirische resultaten van de kwantitatieve studies als de kwalitatieve analyses aangetoond dat ondertitels het product zijn van een delicate evenwichtsoefening waarbij met verschillende, en soms contradictorische criteria, rekening moet worden gehouden. Enerzijds moeten ondertitelaars de richtlijnen van het taalbeleid van de VRT volgen, die voornamelijk de algemene verstaanbaarheid van de televisieprogramma's willen garanderen. Anderzijds moeten ze ook tegemoetkomen aan de noden en wensen van de kijkers, die vaak tegenovergestelde verwachtingen hebben. Bovendien hebben ook de authenticiteit van het programma en technische criteria zoals het beperkte aantal tekens en de tijdsdruk een impact op de taalkeuzes van de ondertitelaars.

Publications

Prieels, L. & De Sutter G. (accepted for publication). 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. *Taal & Tongval.*

Prieels, L. (2017). Balanceren tussen gesproken en geschreven taal : over taalvariatie en taalpolitiek in de ondertitelingspraktijk. In G. De Sutter (Ed.), *De vele gezichten van het Nederlands in Vlaanderen : een inleiding tot de variatietaalkunde* (pp. 335–353). Gent: Acco.

Prieels, L., & De Sutter, G. (2017). Between language policy and language reality : a corpus-based multivariate study of the interlingual and intralingual subtitling practice in Flanders. *PERSPECTIVES-STUDIES IN TRANSLATOLOGY*, 1-22.

Prieels, L., Delaere, I., Plevoets, K., & De Sutter, G. (2015). A corpus-based multivariate analysis of linguistic norm-adherence in audiovisual and written translation. *ACROSS LANGUAGES AND CULTURES*, *16*(2), 209–231.

De Sutter, G., Plevoets, K., Delaere, I., Willems, A., Vandevoorde, L., & Prieels, L. (2014). Linguïstisch conformisme in vertalingen: een corpusgebaseerd multivariaat onderzoek naar lexicale variatie in het Belgische Nederlands. *Beschouwingen uit een talenhuis : opstellen over onderwijs en onderzoek in de vakgroep Vertalen, Tolken en Communicatie aangeboden aan Rita Godyns* (pp. 205–209). Gent: Academia Press.

Appendix

Appendix 1. Overview of Creeber's (2008) genre classification

Genre	Subcategory
Drama	The Single Play, The Western, The Action Series, The Crime Series, Hospital Drama, Science Fiction, Drama-Documentary, The Mini-Series, Costume Drama, The Teen Series, Postmodern Drama
Soap Opera	Soaps, The Telenovela
Comedy	Sketch Comedy, Situation Comedy
Children's television	Children's Drama, Children's Factual, Children's Entertainment, Children's Pre-school (1-5 years), Children's Animation
News	Children's news, Citizen Journalism
Documentary	Observational Documentary, Educational Programming
Reality TV	Docusoaps, Reality Talent Shows, Makeover Shows (Lifestyle), Game shows
Animation	Children's Cartoons, Adult Animation
Popular Entertainment	The Quiz Show, The Game Show, The Celebrity Talk Show, The Confessional Talk Show, Sport, Music on Television, Ordinary Television, Daytime TV, Advertising

Appendix 2. Explanation of the parameters and values of the genre classification in case studies 2 and 3

Parameter/Value	Explanation
Main purpose	What does the program (want to) achieve?
Target audience	A particular group at which a product such as a film or advertisement is aimed (Oxford Dictionaries)
Cast	The actors taking part in a play, film, or other production (Oxford Dictionaries)
Entertaining	Providing (someone) with amusement or enjoyment (Oxford Dictionaries)
Laughing	Making the spontaneous sounds and movements of the face and body that are the instinctive expressions of lively amusement and sometimes also of derision (Oxford Dictionaries)
Informing	Giving (someone) facts or information (Oxford Dictionaries)
Infotaining	Aiming to both to entertain and inform (Oxford Dictionaries)
All ages	Viewers of all ages (4+)
Children	Viewers between 4 and 13 years old.
Actors	Characters who are directly involved in the making of the TV program and who play a role by using a script (real actors, but also presenters and the voice over)
Non-actors	Interviewees, people that are not professional actors or who have nothing to do with the making of the TV program.

Appendix 3. The total number of BSD and CBD attestations per data set (case study 2)

Dataset	Label	Program genre									
		Children's TV		Documentaries		Fiction		Light entertainment		Comedy	
		n	%	n	%	n	%	n	%	n	%
Lexical profiles	BSD	721	10.2	1025	14.5	2236	31.7	2757	29.1	314	4.5
	CBD	72	2.9	89	3.6	1785	72.3	343	13.9	179	7.3
Constructional profiles	BSD	1573	9.2	3673	21.5	5999	35.1	5354	31.3	503	2.9
	CBD	503	9.0	534	9.5	2703	48.2	1683	30.0	183	3.3
Syntagmatic profiles	BSD	234	7.9	811	27.4	1021	34.6	797	27.0	92	3.1
	CBD	83	6.2	231	17.3	561	42.0	395	29.6	65	4.9

Dataset	Label	Program purpose							
		informing		infotaining		entertaining		laughing	
		n	%	n	%	n	%	n	%
Lexical profiles	BSD	1540	21.8	1931	27.4	3030	43.0	552	0.8
	CBD	113	4.6	242	9.8	1431	58.0	682	27.6
Constructional profiles	BSD	4853	28.4	4246	24.8	6613	38.7	1390	8.1
	CBD	988	17.6	1308	23.3	2733	48.8	577	10.3

Syntagmatic profiles	BSD	1024	34.7	607	20.5	1139	35.5	185	6.3
	CBD	308	23.1	313	23.4	613	45.9	101	7.6

Dataset	Label	Target audience			
		all ages		children	
		n	%	*n*	%
Lexical profiles	BSD	6332	89.8	721	10.2
	CBD	2396	97.1	72	2.9
Constructional profiles	BSD	15529	90.8	1573	9.2
	CBD	5103	91.0	503	9.0
Syntagmatic profiles	BSD	2721	92.1	234	7.9
	CBD	1252	93.8	83	6.2

Dataset	Label	Cast			
		actors		non-actors	
		n	%	*n*	%
Lexical profiles	BSD	2756	39.1	4297	60.9
	CBD	2012	81.5	456	18.5
Constructional profiles	BSD	6895	40.3	10207	59.7
	CBD	2935	52.4	2671	47.6
Syntagmatic profiles	BSD	1134	38.4	1821	61.6
	CBD	632	47.3	703	52.7

Dataset	Label	Source language					
		inter<EN		intra.BD		intra.ND	
		n	%	*n*	%	*n*	%
Lexical profiles	BSD	640	9.1	6349	90.0	64	0.9
	CBD	6	0.2	2459	99.6	3	0.2
Constructional profiles	BSD	3012	17.6	13978	81.7	112	0.7
	CBD	232	4.1	5345	95.3	29	0.6
Syntagmatic profiles	BSD	550	18.6	2377	80.4	28	1.0
	CBD	84	6.3	1251	93.7	0	0.0