A functional analysis of the linguistic variation in Flemish spoken commercials

Sofie Van Gijsel, Dirk Geeraerts, Dirk Speelman

KU Leuven – RU Quantitative Lexicology & Variational Linguistics
Blijde Inkomststraat, 21 – 3000 Leuven – Belgium
sofie.vangijsel@arts.kuleuven.ac.be, dirk.speelman@arts.kuleuven.ac.be

Abstract

This article discusses a statistical analysis of the language use of a corpus of spoken commercials, broadcast on Flemish radio and television. Advertising typically attempts to persuade the consumer's public by evoking an identification process with the product promoted. To optimize this identification process, copywriters exploit a number of social trends and evolutions, an important part of which is the linguistic register. In this sense, it can be assumed that the commercials form an ideal corpus to study the linguistic variation –and the linguistic attitudes- in present-day Flanders.

The language use in Flanders is characterized by a stratified continuum in which the informal, substandard register of Dutch, termed '*tussentaal'* (literally: 'in-between language'), takes up an (increasingly) central position, between the regional dialects and standard Dutch. This research will focus on the presence of *tussentaal* characteristics in the corpus. To explore the functional motivations which account for the register variation encountered, a regression analysis is carried out, using a number of dependent variables such as the medium, the format and the target group of the commercial.

Keywords: language variation, regression analysis, spoken advertising language, Dutch.

1. Introduction

In this article, the results of a linguistic analysis of a sample of television and radio commercials, broadcast on Flemish channels, are discussed. As will be explained, one of the essential goals of advertising is getting the potential consumer to identify with the advertising message and the product presented. An efficient technique to stimulate this identification process is the use of an appropriate linguistic register. This entails that the corpus chosen provides trend-sensitive linguistic material, offering insights in the language situation in present-day Flanders.

The Flemish language situation is characterized by a stratified continuum of linguistic registers: although a Flemish version of standard Dutch exists, a substandard, yet not dialectal linguistic variant, dubbed '*tussentaal'* (literally 'in-between language'), is often used, especially in less formal situations. The analysis in this article will especially focus on *tussentaal*, recording (the frequencies of) the different linguistic characteristics in the corpus. The correlations between the variation encountered and a number of extralinguistic factors (e.g. the spot format) will be studied in a regression analysis. Thus, this statistical analysis offers insight in the *functional* motivations which account for the *register variation* in the commercials.

The organization of this article is as follows. First, a short introduction to the field of advertising is given, showing the importance of the linguistic register in commercials. A brief

overview of advertising language research will point at the differences with the proposed analysis. The following section discusses the current language situation in Flanders in some more detail, focusing on the position of *tussentaal*. In the next section, the material and method are presented, followed by the analysis and a discussion of the results. Finally, a brief conclusion is given.

2. The language of advertising

Advertising forms part of the 'marketing (communication) mix' (Borden, 1967), a number of techniques aiming at an optimal communication between marketers and the target public. In a number of ways, advertising differs from a prototypical communication situation (Batra *et al.*, 1996): not only is the coding process complicated, since the sender is a group of *spokes-persons* rather than a single person, but the decoding process is also exceptional in that the receiver, who is part of the consumer public, is unable to provide (direct) feedback. As a consequence, in order to create efficient advertising, marketing research aims to analyze how the consumer *processes* commercials. The efficiency of a commercial is enhanced if the consumer *identifies* with the advertising message, since this creates a positive attitude towards the product advertised. This identification process can be brought about by developing a good product *image* (or *brand personality*) or by creating *group norms*, which suggest that certain people or spokespersons, with whom the target public identifies, positively validate the product. Fauconnier and Van der Meiden (1993: 21) point out that in this process, a number of cultural and social factors play a role, an important one of which is the *linguistic register* used.

Broadly speaking, the language of advertising has been analyzed from two perspectives. Some research concentrates on the persuasive –and often misleading- effects of advertising. Geis (1982) for example shows how advertising makes use of techniques such as conversational and pragmatic implicatures, ellipses and hedges, which make the consumer assume stronger claims than are acceptable from a strictly logical point of view. In the same vein, Vestergaard and Schröder (1985) analyze ads in magazines, discussing a number of (rhetoric) devices which exploit cultural presuppositions and images in order to 'lure' the public into purchasing the presented products. Similarly, Goddard (1998) and Cook (2001) study the discourse of advertising, focusing on the interaction between textual (e.g. connotations) and contextual elements (e.g. the music and/or the pictures or the 'narrative voice' of the speaker).

Another strand of research studies the advertising genre from a more linguistic-technical point of view, mostly focusing on commercials in a single language. For example, research analyzing Dutch advertising language includes Lemmens (1994) and Gijsbers *et al.* (1998), who study the use of English in Dutch commercials, while Stroobants (1997) studies the linguistic differences between commercials broadcast in The Netherlands and in Flanders (e.g. the word order of the verbal group).

In this research, the aim is to combine the two perspectives. In an empirical, corpus-based analysis, the linguistic characteristics of the commercials will be analyzed. The variation encountered will be related to a number of extralinguistic factors which *functionally motivate* the register chosen in the advertisement corpus.

3. The present language situation in Flanders

After a century-long struggle for a Dutch standard language in Flanders (see for example Van den Toorn *et al.*, 1997), a Flemish norm for standard Dutch has emerged,[1] which differs from standard Dutch as spoken in The Netherlands in a number of phonetic, morphological, lexicological, and grammatical aspects (see Geeraerts *et al.*, 1999 for an overview). Yet, in more informal strata of Flemish Dutch, the standardization process seems to have halted, giving way to the use of an intermediate, substandard register of Dutch, dubbed *'tussentaal'* (literally '*in-between* language') or '*Verkavelingsvlaams'* (Van Istendael, 1989). This register, which is positioned on the linguistic continuum between the regional dialects and standard Dutch, seems to be a more 'natural' language for many Flemish, while the standard language is often considered an external, artificial linguistic norm (e.g. Jaspaert, 1986; Geeraerts, 1998). Although *tussentaal* functions *de facto* as an informal vernacular on a supraregional level, it is characterized by a number of *substandard* features, especially on the level of pronunciation and morphology. Geeraerts *et al.* (1999) have also shown that the distance between *tussentaal* and Belgian standard Dutch on the one hand is larger than that between the Netherlandic informal register and standard language on the other hand.

It is largely due to its rather recent emergence and substandard character that relatively little research has focused on the (possibly) systematic linguistic characteristics of *tussentaal,* although Smedts (2000) and Geeraerts (2001) present an overview of *tussentaal*-features. In the analysis presented here, the substandard linguistic features occurring in the supraregional corpus of spoken commercials will be identified (see 4.2.2 for an enumeration of the different *tussentaal*-characteristics analyzed). To explore what functional motivations account for the use of *tussentaal*, the linguistic register is correlated to a number of extralinguistic factors.

4. Material and method

4.1. *Material*

The sample of commercials analyzed consists of 150 radio and 150 television ads, broadcast on Flemish, supraregional channels.[2] All material was collected during the Fall of 2000. In order to collect a representative sample, commercials from the different Flemish channels and for different times of the day were recorded. Every commercial was included only once in the sample.

4.2. *Method*

Based on research as presented in 3., the linguistic characteristics encountered in the corpus are identified as either *tussentaal* or standard language. Next, the effect of a number of extralinguistic, independent factors on the language used in commercials is analyzed in a regression analysis. In Section 5, this statistical analysis will be discussed; here, the independent and dependent variables of the statistical analysis are introduced.

[1] This standard language is often dubbed '*VRT-Dutch*', after the national television broadcast channel, VRT, whose news reporters are generally perceived as providing a good example of 'standard Dutch' as it is spoken in Flanders.

[2] For radio ads, material was taped from the Radio 1 and Radio 2, Studio Brussel, Donna and Klara, while for the television commercials, the national channels *VRT* and *Canvas* and the commercial channels *VTM*, *Ka2* and *VT4* were taken into account.

4.2.1. *The independent variables*

A first independent variable which is taken into account is the *medium* of the ad: it can be expected that the linguistic register of the commercial depends on whether it is broadcast on radio or on television. While television ads make use of a number of visual aids to convince the potential consumer, the techniques available for radio commercials are more limited (see also Hagerman, 1990: 159). Therefore, for radio ads, a more creative use of the linguistic register might be expected.

Secondly, the *target public* was a potentially significant factor: is it possible to find register differences in commercials specifically aiming at a public of children (*Child*), adolescents (*Adolesc*) or a more general (mostly adult) public (*Gen*)?[3]

A third factor analyzed is the *spot format*. As Hagerman (1990: 125) explains, a format is 'a combination of structure and content that acts as a catalyst to hold the listener's interest'. Although in works on advertising techniques, the exact number and types of formats are slightly different (see for example Baldwin, 1989 or Batra *et al.*, 1996), for this research, it was sufficient to work with relatively general and easily discernable categories. Broadly speaking, two main types of formats can be distinguished. On the one hand, a number of formats adopt an informative, *rational-argumentative* approach, by focusing on the product itself (so-called *product presentations*) or by showing its use (*demonstration*-spots). Other formats, on the other hand, are more *emotional-suggestive,* in that they attempt to persuade the public by evoking (positive) feelings. An important format type belonging to this second category is the *minidrama*, in which actors play a certain scene (for example, a *slice-of-life* minidrama for cereals could show a family having breakfast). Other formats include *personality* spots, where a character or (cartoon) figure is linked to the product (e.g. *Tony the Tiger* for *Frosties* of Kellog's) and *testimonials*, in which an expert, 'the man on the street' or a celebrity testifies about the product. Minidrama's differ from the other emotional-suggestive formats in that most often, more than one actor is involved, interacting in dialogic speech. Since this dialogic aspect might have an influence on the linguistic register, this spot format is analyzed separately. Finally, each of the different formats discussed can be followed by a *pay off*, a short format element which is typically read by a *voice-over*, who repeats the brand name and/or a slogan. Since it is interesting to test if the linguistic register of the voice-over is different from the language in the spot body, the *format elements* rather than the entire spots are taken as observations in the regression analysis. Thus, each of the format elements of the corpus was assigned to one of the following categories: 1) rational-argumentative spots (*Rat*), 2) minidrama (*Md*), 3) emotional-suggestive spots (*Em*) and 4) pay off *(Po)*.[4]

A last factor analyzed was the *spot subject*, according to 7 categories (toys, technology & communication, cars, financial services, media & culture, products for health and personal care and non-profit ads[5]). Since none of the spot subjects showed a significant relation with the linguistic register, this factor will not be discussed in detail.

[3] Since occasionally, it was rather difficult to draw the line between 'adolescent' and 'adult' spots, only those ads which could clearly be distinguished (for example based on the spokespersons or the product advertised) were classified as adolescent or *Adolesc*-spots.

[4] Of course, the division between the different format types is not always clear-cut; indeed, efficient commercials will often combine different persuasion techniques. Yet, only in very few cases, it was doubtful which format type an element had to be assigned to.

[5] Non-profit spots are sponsored by the government or non-profit organisations (e.g. for a drink-drive campaign) and most often broadcasted on national channels. Therefore, they could be expected to have a higher use

4.2.2. *The dependent variable*

As was mentioned before, the analysis presented focuses on the stratified continuum of spoken language use in Flanders. Thus, a number of spot elements in the corpus, which made use of linguistic registers falling outside this continuum, were consequently excluded from the dataset used in the regression analysis. These were English-spoken elements, a small number of television spots in which no spoken language was used, one purely dialectal spot element and a few spot elements in which Netherlandic Dutch was spoken. Globally speaking, 409 spot elements were analyzed, while 53 were left out.

Instead of working with a binary classification of the spot elements in *tussentaal* and standard language, a *tussentaal*-index was calculated for each of these 409 elements, based on the relative frequency of all *tussentaal*-characteristics vs. all linguistic characteristics encountered in the spot element. This way, the intrinsic continuum of the linguistic register is taken into account. The 17 linguistic characteristics analyzed are the following:[6]

Phonological characteristics:

- word-final t-deletion in words like 'goed' ('good'), 'niet' ('not'), 'wat' ('wat') and 'dat' ('that'), 'met' ('with') and end deletion in 'maar' ('but')
- regional accent of Brabant (the central province of Belgium)
- word-initial h-dropping

Morphological characteristics:

- variants of the personal pronouns of the 1st person singular (*(e)kik* instead of 'ik'), the 2nd person singular (*gij* instead of 'jij'/'you' and the 3rd person singular (*hem* instead of 'hij'/'he')
- variants of the attributive adjective when followed by a *d, t, b, h* or vowel
- variant of definite article 'de' ('the'): *den* or *nen*
- variant of the indefinite article 'een' ('a'): *ne* or *e*
- variants of the demonstrative pronoun (*dieje/diene/dezen* for 'die/deze' or 'these/those'), possessive pronoun (e.g. 'mijn*en*' for 'mijn'/'mine') and *een/geen* ('one/none'), if followed by a *d, t, b, h* or vowel
- variant of the diminutive: *–(s)ke* instead of '–je' (and its allomorphs)
- variant of imperative singular: *root+t* instead of *root*

5. Linear regression

As mentioned before, a *linear regression* analyzes the impact of the independent factors (in this case medium, format type, target group and subject) on the dependent factor, the linguistic register used in the spot elements.[7] The statistical model is built using a *forward stepwise procedure*: the independent variables are added to the model in their order of significance; if

of the standard language. Although a slightly higher use of standard Dutch was found, the result was not significant.

[6] Although we find a few other morphological variants (such as plural pronouns), some minor *syntactic* deviations (e.g. in the order of the verbal group or the double negation) and *lexical* differences (e.g. substantives and interjections), these elements were extremely infrequent in the corpus. Therefore, they were not analyzed in further detail.

[7] In *Appendix A*, a table is included which gives an overview of the absolute numbers of the independent variables in the entire set of 409 observations, split up for '*tussentaal'* and 'standard language'. For ease of representation, in this table, each spot element containing one or more *tussentaal*-characteristics was classified as '*tussentaal'*. Of course, it should be kept in mind that, since the linguistic registers form a continuum, the boundary between the *tussentaal* and standard Dutch is rather artificial; in the regression analysis, a *tussentaal-index* is used rather than a binary classification to account for this gradualism.

the next variable to be added to the model does not contribute significantly to the proportion of explained variation, the analysis automatically stops. Not only will it be possible to test if all the extralinguistic factors analyzed have a statistically significant effect on the linguistic register, but also, the statistical output indicates the *effect of the individual variables* by way of the *regression coefficient*. This coefficient, also called the *B-value* or *estimate* of the independent variable, gives a measure of how the dependent variable (in our case, the *tussentaal*-index) increases if the other variables in the model are kept constant. Further, it is possible to evaluate the *global model* by interpreting the *R-squared value*: this value between 0 and 1 indicates how much of the variation in the data is explained by the model. Finally, a *p-value* for the significance of the global model is also given.

As was mentioned before, the results for 'subject' were not significant. Therefore, it comes as no surprise that the stepwise procedure does not include this factor in the model. The other three factors significantly contribute to the explained variation, with 'format' being added first to the model, followed by 'target group' and 'medium'. An overview of the p-values of the different factors, and the estimates (or B-values) with their confidence intervals (C.I.) is given in *Table 1*:[8]

	p-value	*Estimate*	*C.I.*
Format: minidrama	< 0.001	0.312	0.270-0.352
Public: adolescents	< 0.001	0.225	0.118-0.331
Medium: television	0.0125	-0.044	-0.079-(-0.009)

Table 1. The P-values, estimates and respective confidence intervals for the significant independent variables influencing the tussentaal-*index of the spot elements*

It is clear that the factor 'minidrama' is highly significant ($p < 0.001$), with the estimate indicating that the model gives this format element 31.2% more for the *tussentaal*-index than to the other format elements. For the factor 'target public', the *estimate* is 0.225 for the value 'adolescents', which shows that *tussentaal* is predicted to be 22.5 %higher in *adolescent*-spot elements than in those having children or adults as main target audience. Admittedly, the confidence interval is rather large here ([0.118-0.331]), due to the small number of *Adolesc*-elements in the sample (see Appendix A). The third most important factor influencing the linguistic register is the *medium*, with 'television' predicting the *tussentaal*-index to decrease with 5 %. Here, the confidence interval is much smaller ([-0.079-(-0.009)]), due to the large number of observations (162 television-formatelements).

Next, it is possible to evaluate the *global model*, which takes up 'format element', 'target group' and 'medium' as factors in the regression. The model is highly significant ($p < 0.001$), which means that the chance that the *tussentaal*-index is not influenced by the independent factors (as the null hypothesis would claim) is virtually non-existent. The *R-squared value* is 0.492, which means that the model explains about 50 %of the variation in the data. Therefore, it can be concluded that the model predicts the degree of *tussentaal* in the spotelements fairly well.

[8] The values of the confidence interval indicate that there is a 95% chance that the real value, for which the B-value is an estimate, lies between these values. The confidence interval will typically be larger if the estimate is based on relatively few observations; this should be taken into account when interpreting this measure.

The results show that the use of *tussentaal* in the corpus of Flemish spoken commercials is strongly influenced by a number of independent factors. A first important factor is the *spot format*: in dialogic *minidramas*, *tussentaal* is used extensively. Most likely, this informal register is used to make the spot sound more 'natural' and 'authentic'. The recurrent combination of a *tussentaal*-minidrama with a pay off in which standard Dutch is spoken shows how the *style-shifting* is used consciously and strategically to initiate and mark a situational change (see e.g. Milroy and Gordon, 2003 for a discussion of the notion of 'initiative style-shift'). Secondly, in radio commercials, *tussentaal* is used more often than in television spots. This may well be due to the fact that radio spots can only make use of sound elements to persuade the consumer, while television ads can exploit a wide array of (audio)visual methods to evoke an identification process. Finally, it is important that the *tussentaal*-index is higher in spots aiming at a public of adolescents. As Bell (2001) and Milroy and Gordon (2003: 207) explain, in order to express identification with particular groups (which may well be geographically distant, as is the audience of commercials), speakers can exploit available linguistic resources. Here, the *adolescent*-audience brings about a strategic style-shift towards a less standardized register. Therefore, it might be hypothized that *tussentaal* has young, even somewhat rebellious, connotations, as opposed to the 'conformist' norm of the standard language.

These conclusions tie in with independent linguistic research (e.g. Jaspaert, 1986; Deprez, 1994; Geeraerts, 1998), which shows that many Flemings consider the standard language as an artificial, external norm rather than an authentic or natural register, especially in slightly less formal, spoken language. Also, the connotations which seem to be associated with *tussentaal* are in line with what Geeraerts (2001) shows for the language used in Flemish soaps: *tussentaal* is often used when a natural, informal register is required to ensure easy identification by the spectators.[9] It seems to be the case that in the advertisement corpus analyzed, the 'authentic', natural-sounding register *tussentaal* is equally exploited as an efficient technique to enhance an identification process of the consumer with the product.

6. Conclusion

In this article, the analysis of a corpus of spoken commercials broadcast on Flemish radio and television was discussed. The advertisements provided an interesting view on the current language situation in Flanders, which is characterized by a stratified continuum of linguistic variants. More specifically, the use of the 'in-between' register *tussentaal*, filling the gap between 'dialects' and 'standard Dutch', was studied in detail. In order to find out what *functional* motivations lie behind the language variation used in the spots, a regression analysis was carried out, investigating the influence of four independent factors on the linguistic register. The results show that radio (as opposed to television) spots, an adolescent target public and dialogic minidramas significantly increase the *tussentaal*-index. These spot types all attempt to sound natural, informal and realistic; connotations which typically seem to be associated with the register *tussentaal*.

[9] Binnemans (2003) comes to similar conclusions in her analysis of the Flemish television travel program '2x Enkel': while the young, somewhat 'wild' male reporter of the program uses *tussentaal* quite intensively, in order to sound natural, casual and a bit rebellious, the older, female reporter, who often presents the more informative sections of the program, attempts to use a more formal, standardized linguistic register.

References

Baldwin H. (1989). *How to Create Effective TV Commercials.* NTC Business Books.

Batra R., Myers J.G. and Aaker D.A. (1996). *Advertising Management.* Prentice-Hall International.

Bell A. (1999). Styling the other to define the self. *Journal of Sociolinguistics*, vol. (4): 523-541.

Binnemans T. (2003). *Taal en Identiteit: Een Functionele Analyse van de Taalvariatie in het Televisieprogramma* 2x Enkel. Unpublished dissertation, University of Leuven.

Cook G. (2001). *The Discourse of Advertising (2nd edition)*. Routledge.

Deprez K. (1994). Waarom ze hun Vlaams liever zien verdwijnen. *Onze Taal,* vol. (62): 251-253.

Fauconnier G. and Van der Meiden A. (1993). *Reclame. Een Andere Kijk op een Merkwaardig Maatschappelijk Fenomeen*. Dick Coutinho.

Geeraerts D. (1998). VRT-Nederlands en Soap-Vlaams. *Nederlands van Nu*, vol. (3): 75-77.

Geeraerts D., Grondelaers S. and Speelman D. (1999). *Convergentie en Divergentie in de Nederlandse Woordenschat. Een Onderzoek naar Kleding- en Voetbaltermen*. Meertens Instituut.

Geeraerts D. (2001). Everyday language in the media. The case of Belgian Dutch soap series. In Kammerer M., Konerding K., Lehr A., Storrer A., Thimm C. and Wolski W. (Eds), *Sprache im Alltag. Beitrge zu neuen Perspektiven in der Linguistik Herbert Ernst Wiegand zum 65. Geburtstag gewidmet*. Walter de Gruyter: 281-291.

Geis M. (1982). *The Language of Advertising.* Academic Press.

Gijsbers I., Gerritsen M., Korzilius H. and van Meurs F. (1998). Engels in Nederlandse tv-reclame. *Onze Taal*, vol. (7/8): 175-177.

Goddard A. (1998). *The Language of Advertising: Written Texts*. Routledge.

Goossens J. (2000). De toekomst van het Nederlands in Vlaanderen. *Ons Erfdeel*, vol. (43): 3-13.

Hagerman W.L. (1990). *Broadcast Advertising Copywriting.* Stoneham. Butterworth Publishers.

Jaspaert K. (1986). *Statuut en Structuur van Standaardtalig Vlaanderen*. Leuven Universitaire Pers.

Lemmens M. (1994). Amazing…moderne reclametaal. *Onze Taal*, vol. (7/8): 176-177.

Milroy L. and Gordon M. (2003). *Sociolinguistics: Method and Interpretation*. Blackwell.

Rietveld T. and van Hout R. (1993). *Statistical Techniques for the Study of Language and Language Behaviour*. Mouton de Gruyter.

Smedts W. (2000). *Project Taalhantering VRT 1999-2000. Eindevaluatie.* VRT-project Taalhantering, Internal Report, University of Leuven.

Stroobants K. (1997). Noord/Zuid-verschillen in de taal van de reclame. *Taal en Tongval*, vol. (10): 140-151.

Van den Toorn M.C. (Ed.) (1997). *Geschiedenis van de Nederlandse Taal*. University Press.

Van de Velde H. (2000). 100 jaar uitspraak. *Nederlands van Nu*, vol. (3): 35-38.

Vestergaard T. and Schrøder K. (1985). *The Language of Advertising.* Basil Blackwell.

Van Istendael G. (1989). *Het Belgisch labyrint*. *De schoonheid der wanstaligheid*. Arbeiderspers.

Appendix

		Standard language (306)			*Tussentaal (103)*			*Total (409)*
		Child	Adolesc	Gen	Child	Adolesc	Gen	
Radio		3	2	164	0	7	71	247
	Rat	2	1	47	0	0	2	52
	Md	0	0	18	0	4	67	89
	Em	1	0	2	0	0	1	4
	Po	0	1	97	0	3	1	102
TV		28	3	106	5	0	20	162
	Rat	17	1	49	0	0	0	67
	Md	3	0	12	4	0	17	36
	Em	0	0	11	1	0	2	14
	Po	8	2	34	0	0	1	45
		31	5	270	5	7	91	**409**

Table 2. The absolute numbers of the independent variables 'target public', 'format' and 'medium' for the 409 format elements, split up for 'standard language' and 'tussentaal'

Legend:

Target public:
- Child: children
- Adolesc: adolescents
- Gen: general

Format:
- Rat: rational-argumentative
- Md: minidrama
- Em: emotional-suggestive
- Po: pay off

Medium:
- Radio
- TV