KU LEUVEN
FACULTY OF ARTS
DEPARTMENT OF LINGUISTICS

New approaches to measuring the social meaning of language variation

Exploring the Personalized Implicit Association Test and the Relational Responding Task

Thesis presented in partial fulfilment of the requirements
for the degree of Doctor in Linguistics

by Laura Rosseel

Supervisor: Prof Dr Dirk Geeraerts
Co-supervisor: Prof Dr Dirk Speelman

Leuven 2017

Acknowledgements

There are many people I wish to express my gratitude to for their role in the realisation of this thesis. First and foremost, I want to thank my supervisors Dirk Geeraerts and Dirk Speelman, for the inspiring discussions and invaluable feedback, for four years of encouragement and advice, and for challenging me and guiding me on my way towards a PhD. Their door was always open and they never failed to make time to talk with me about any ideas or doubts I had. I could not have asked for a better team of supervisors.

I am also grateful to Carmen Llamas, Stefania Marzo, Robert McKenzie and Dorien Van De Mieroop for accepting the invitation to be part of my examination committee, and to Hubert Cuyckens for acting as the chair of the committee. Dorien Van De Mieroop was also a member of my supervisory commission together with Adriaan Spruyt. I would like to thank them both for the interest they have shown in my work and for their feedback from various perspectives. I owe Adriaan much gratitude as well for introducing me to the field of social psychology and for his enthusiastic methodological support.

I am much indebted to the Research Foundation Flanders (FWO) for granting me a PhD fellowship which allowed me to work on my research independently for four years. I am also thankful for the many respondents who were kind enough to sacrifice some of their time to take part in my experiments, and to the colleagues in Leuven, Kortrijk and Antwerp who assisted me with the practical side of administering the experiments. In particular, I want to thank the researchers of the CLiPS group at the University of Antwerp for lending me their experimental infrastructure and to José Tummers and his

Limburgish colleagues at UCLL for allowing me to use their recording studio, as well as their voices.

I have had the privilege to work on this thesis in a fantastic research group. I wish to thank all QLVL members for their professional input, but also for their support and friendship. Eline and Stefania, thank you for your advice on so many things; and Jason, thank you for being a great office mate. To Alek, Benedikt, Dirk, Eline, Freek, Isabeau, Jason, Jeroen, Jocelyne, Karlien, Kris, Kristina, Leonie, Melanie, Stefano, Thomas – the QLVL lunch crew: thank you for the most intriguing conversations during lunch and coffee breaks. That also goes for Charlotte, Dana, Lena and the many colleagues from other research units at the faculty who have made my years as a PhD student in Leuven an unforgettable experience. A word of thanks is also in place for Sylvia and Sonja, who provided kind and reliable administrative support. Additionally, I am indebted to many colleagues from various universities for their feedback and interest during conferences and especially to Stef Grondelaers for organising a workshop on implicitness in language attitude research with me.

A final word of thanks goes to my friends, to my family and to Daniel. Cheers to my friends for their encouragement and interest in my work as a linguist, in particular to those who willingly served as guinea pigs during the preparation of experiments, and to Laura Bergans for designing the cover of this book. A special mention is due for my sister Elisabeth, the head guinea pig, who was always prepared to give honest feedback on early versions of experiments from the perspective of a participant. I am also greatly indebted to Serina Hayes who has been kind enough to proofread my texts. All remaining errors are of course my own. Furthermore, I want to thank my grandparents for their words of wisdom and their faith in me. I also wish to use this opportunity to express my warmest gratitude to my parents. They have supported me unconditionally in my studies and in everything else I have undertaken. And finally, Dan, thank you so much for standing by me and being my most loyal supporter.

Laura

Leuven, July 2017

Table of contents

List of tables

List of figures

Introduction

The objective of this work is twofold. It sets out to contribute to the study of language attitudes on a methodological as well as a descriptive level. The main goal of the study is the methodological one. Notwithstanding some exceptions (e.g. Preston 1982), quantitative language attitude research has known little methodological innovation since the introduction of the matched-guise technique in the 1960s (Lambert et al. 1960). This relative methodological stagnation forms a stark contrast with the recent explosion of new methods in social psychology. The latter field is heavily invested in the study of attitudes and has witnessed the development of a large amount of new attitude measures in the last few decades (Petty et al. 2009; Gawronski & De Houwer 2014). In this thesis, we investigate whether linguistic attitude research can take advantage of these recent developments in psychological attitude research: is it possible to use those new attitude measures to study the social meaning of language variation?

The second goal of our research is a descriptive one. In the past decades, attitudes towards Dutch language variation have been relatively understudied. In the Netherlands, this situation is changing as more linguistic attitude studies are being carried out (cf. recent work by Grondelaers et al. 2009; Grondelaers & Van Hout 2010). By contrast, in Flanders, the Dutch speaking part of Belgium, there is still much work to be done. Against this background, we aim to make a modest contribution to the emerging picture of the language attitudinal landscape in Flanders.

To accomplish these goals, three studies were conducted. These studies focus on three different social psychological attitude measures and feature five varieties of Belgian Dutch. The first study explores the Personalized Implicit Association Test (P-IAT, Olson & Fazio 2004) to measure attitudes towards Standard Belgian Dutch and two regional varieties of Belgian Dutch (Chapter 2). The P-IAT is compared to two other attitude measures in this study: a direct rating task and a second social psychological technique called affective priming (AP, Speelman et al. 2013). Following up on the successful implementation of the P-IAT as a language attitude measure in the first study, the second study continues the exploration of the P-IAT (Chapter 3). Experimental methods like the P-IAT often decontextualise the attitude object. Given the abundant evidence in both the social psychological and sociolinguistic literature that context significantly influences attitudes (e.g. Gawronski & De Houwer 2014; Campbell-Kibler 2007; Soukup 2013a), we set out to investigate whether we could introduce context cues in the design of the P-IAT. If successful, the method could be used to measure attitudes in a more contextualised way. In the final study, which is reported in Chapter 4, we investigate the potential of the Relational Responding Task (RRT, De Houwer et al. 2015) as a measure for language attitudes.

To provide the necessary background for the studies reported in Chapters 2-4, we first present a state of the art of attitude measurement in linguistics and social psychology (Chapter 1). In addition to this overview, we will discuss in detail how the social psychological measures under study (AP, the P-IAT and the RRT) function. This is followed by a comparative review of their procedural characteristics, which points out potential advantages and shortcomings in the light of their use in sociolinguistic attitude research.

The final chapter of this thesis (Chapter 5) summarises the outcome of the three studies and points to questions that remain unanswered, but most importantly, it situates our work in a larger framework of research to illustrate its theoretical relevance. Although the main objective of this thesis is to provide novel tools for sociolinguistic research on language attitudes, it also goes beyond this methodological aim. That is to say, our work fits within the tradition of Cognitive Sociolinguistics. This emerging field aims to combine the

strengths of sociolinguistics and Cognitive Linguistics, namely the empirical study of intralinguistic variation and the study of meaning (Geeraerts & Kristiansen 2015). That Cognitive Linguistic project takes shape in this thesis through the study of the social meaning language users attribute to language variation.

Chapter 1

Attitude measurement in linguistics and social psychology: A state of the art

In this chapter, language attitudes are explored from a conceptual as well as a methodological perspective. Both aspects of language attitude research have known a stagnation or have received little attention until quite recently. In Section 1 of this chapter, we will consider how language attitudes are defined by linguists and make a brief comparison with social psychological conceptualisations of attitudes. Section 2 focuses on methods to measure attitudes. First, we discuss the traditional methods employed in linguistic attitude research. Subsequently, the linguistic approach will be confronted with that of social psychology, a field where attitude research is core business. The latter field has not shied away from precisely defining its understanding of the construct attitude and continuously challenges these conceptualisations. After briefly sketching the current methodological dynamics in social psychology, we focus on the three attitude measures that take centre stage in the studies reported in Chapters 2-4: affective priming, the Implicit Association Test and the Relational Responding Task. They are reviewed from a procedural point of view in order to provide linguists with a number of criteria that may assist them

in deciding whether these methods offer potential for their research and if so, which of them is most suitable.

1 Language attitudes: A conceptual perspective

This section outlines how sociolinguists manage the construct 'language attitude' from a conceptual point of view, how this compares to the social psychological interpretation of the construct and, finally, how the work reported in this thesis positions itself in this respect.

It seems that precisely defining language attitudes has not been a primary concern for many linguists investigating them. If the construct is defined at all, this is often done by referring to older psychological research that understands attitudes as evaluations in terms of valence (positive vs. negative) (e.g. Garrett 2010 who cites for instance Thurstone 1931 and Sarnoff 1970). This rather narrow interpretation of the concept attitude forms a sharp contrast with what linguists actually study when they conduct language attitude research. The range of perceptual correlates of linguistic phenomena investigated is quite diverse, yet, within that diversity, a number of recurrent patterns have been detected (Garrett 2010). The most commonly reported pattern pertains to the distinction between competence/status/prestige/ superiority (e.g. intelligence, level of education, wealth) and sociability/ solidarity/warmth/social attractiveness (e.g. friendliness, trustworthiness) (Ryan et al. 1982; Zahn & Hopper 1985; Kristiansen 2010b; Garrett 2010). Additionally, another pattern of attitudes is sometimes distinguished and relates to dynamism (i.e. liveliness, enthusiasm) (Zahn & Hopper 1985; Garrett 2010). These three patterns are referred to as dimensions of language attitudes. Because these recurring dimensions have proven relevant in a multitude of speech communities, many studies have relied on them as a starting point. However, blindly relying on these dimensions poses the risk of overlooking any additional dimensions or aspects of language attitudes that may be relevant in a specific community (Garrett 2010).

As an alternative or in addition to a valence based definition of attitudes, some linguists define language attitudes by referring to their structure, which

they posit as tripartite (e.g. Edwards 1982; Ryan et al. 1982; Bradac et al. 2001; Garrett 2010; Lybaert 2014). According to this definition, language attitudes can have a cognitive, affective and behavioural component. The cognitive component concerns beliefs about the attitude object, while the affective one refers to positive or negative feelings towards that object. The third component, behaviour, links the attitude concept to a tendency to act in a certain way. This last point is controversial. Although the link between attitudes and behaviour is not well-studied in linguistic attitude research (with the exception of some studies like Kristiansen 1997), psychological research has demonstrated that the relationship is highly complex with many potential factors moderating it (Ajzen & Fishbein 2005).

Recently, a number of authors have expressed their dismissal of a narrow definition of language attitudes and have chosen to interpret the construct very broadly. One example can be found in Soukup (2013b: 268) where the author indicates she understands language attitudes as 'the *social meanings* associated with language use and variation' (emphasis in the original). Another case in point is Preston (e.g. Preston 2010; ibid. 2015) who prefers the umbrella term 'language regard' under which he groups attitudes (in a strictly evaluative sense), beliefs (interpreted as estimations of the probability that one's information about a certain object/person is accurate), ideology or any 'other less well-specified conceptual bits' (Preston 2015: 5). Note that concepts like 'belief' and 'ideology', but also 'values' or 'stereotypes' are regularly used by linguists in the context of language attitude research. Yet, just like for the concept attitude itself, their delineation often remains difficult and vague (Soukup 2009). For an attempt at distinguishing between these concepts, we refer the reader to Garrett (2010).

This lack of precise definition is also evident with regard to the distinction between different types of attitudes. Most sociolinguists in attitude research make a distinction between overt and covert language attitudes[1]. Even

[1] Note that different authors use different terms to refer to what seems to be the same distinction, e.g. private – public (e.g. Grondelaers & Speelman 2015), deep – shallow (e.g. Grondelaers & Speelman 2015), conscious – subconscious (e.g. Kristiansen 2009; Preston 2015; Grondelaers & Speelman 2013), implicit – explicit (e.g. Grondelaers &

though these concepts are not usually elaborated on[2], the emphasis seems to be on awareness or consciousness in most studies. This focus on awareness can be traced back to the very beginnings of sociolinguistics (Weinreich et al. 1968) with for instance Labov's (1972) distinction between indicators, markers and stereotypes. Despite this central role in sociolinguistics, awareness has not been studied extensively and the exact meaning and function of concepts like overt vs. covert attitudes are poorly understood (Babel 2016; Campbell-Kibler 2016). This is striking given that many researchers suspect covert attitudes in particular to play a crucial role in language variation and change (e.g. Weinreich et al. 1968; Kristiansen 2010b; Preston 2015)[3]. Recently, however, sociolinguistics seems to be witnessing a change as researchers have begun to tackle the vagueness surrounding the concept and function of awareness. This is exemplified by Babel (2016), a recent volume which is dedicated to the topic. In parallel with this development in sociolinguistics at large, linguistic attitude research has begun to pay more detailed attention to the conceptual distinction between overt and covert attitudes. This appears to be the case particularly in research inspired by recent social psychological methods or theory (e.g. Campbell-Kibler 2012; McKenzie 2015a; Pantos & Perkins 2012; Preston 2015; Preston 2016).

This last observation is perhaps not surprising given the amount of research (both theoretical and empirical) in social psychology that has been dedicated to the distinction between implicit and explicit attitudes. That distinction goes beyond the notion of consciousness: psychologists interpret implicit attitudes in terms of automatic processes as opposed to controlled ones (Gawronski & De Houwer 2014). Automaticity is perceived as comprising multiple features (unintentionality, resource-independence, uncontrollability

---

Van Hout 2010; Preston 2011; Grondelaers & Speelman 2013; Watson & Clarke 2013; McKenzie 2015b).

[2] There are some notable exceptions such as Kristiansen (2010b) who carefully analyses these concepts and compares his interpretation of overt vs. covert attitudes with Labov's, or Preston (1996) who analyses Silverstein's (1981) concept of awareness in the light of his own ideas concerning folk linguistics.

[3] The importance accorded to subconscious language attitudes to explain language variation is not uncontroversial (e.g. Soukup 2013b).

as well as unconsciousness) that need not all be present, but can qualify the way in which the outcome of an attitude measure is considered implicit (De Houwer et al. 2009; De Houwer & Moors 2010; Gawronski & De Houwer 2014; see Section 2.2 for a more detailed discussion). It is not only the difference between implicit and explicit attitudes, but also the concept 'attitude' itself that has been theorized more extensively in social psychology compared to sociolinguistics. The social psychological definition of attitudes has changed over the last century (Fazio & Petty 2008). The tripartite model of attitudes, which defines attitudes as having an affective, a cognitive and a behavioural component and is still referred to regularly in linguistic attitude studies, finds its origins in social psychological attitude research (Katz & Stotland 1959). Yet today, many psychologists have abandoned this tripartite view and define an attitude as 'a person's evaluation of an object – favorability or unfavorability toward the object' (Fazio & Petty 2008: 3). Often, attitudes are also described as associations between an attitude object and its evaluation stored in memory (Fazio & Petty 2008). Although the link between attitudes and memory has been heavily debated, most researchers in social psychology nowadays agree that attitudes are not stable memories, and that both information stored in memory and online processing play a central role (Albarracín et al. 2005).

In the studies reported in Chapters 2-4, we have chosen to conceptualise language attitudes quite broadly as the association between language phenomena and social meaning(s). As a result, the terms *language attitudes* and *social meaning of language (variation)* will be used interchangeably. A broad definition in terms of social meaning is consistent with the interpretation of a number of linguists who have recently argued against a narrow interpretation of attitudes (e.g. Soukup 2013b). In that sense our definition comes close to Preston's (2015) language regard, although we will not go as far as to investigate higher order structures like language ideologies, which are included in language regard. That does not mean, though, that we deny these structures are closely related to attitudes (cf. Garrett 2010). Our broad definition also ties in with the most recent social psychological definitions of attitudes which have moved away from more restricting views such as the tripartite model of attitudes. Defining language attitudes in terms of

associations between language phenomena and social meaning also allows us to connect the field of linguistic attitude research to, and ground it in other movements, theories and ideas in (socio)linguistics. For instance, social meaning is a core concept in third wave sociolinguistics, while thinking about attitudes as associations in memory opens the door to exemplar theory (cf. Chapters 3 and 5).

As for implicit vs. explicit attitudes, we will follow the social psychological interpretation in terms of automaticity rather than consciousness alone. Although we will come back to the psychological vs. linguistic approach to implicitness in Section 2, it may be useful to explain already at this point that we have chosen this more elaborate interpretation for both methodological and theoretical reasons. Firstly, the methods we explore in Chapters 2-4 are grounded in a social psychological definition of implicitness. Secondly, we believe it is worthwhile for linguists to consider an interpretation of attitudes that goes beyond consciousness. Further research will perhaps show that the aspect of implicitness most crucial to models of sociolinguistic cognition is consciousness after all, but without theoretical and empirical investigation we will not find out. Finally, employing a multi-component definition of implicitness offers the possibility to conceptualise implicitness and explicitness as two extremes on a continuum rather than as a binary distinction. A measure can tap into processes which are implicit in certain ways, but not in others. It also makes it possible to qualify a measure as more or less implicit than another measure. This gradient nature of implicitness will be called upon in Chapters 2 and 4.

2 (Language) attitude measurement: A methodological perspective

Language attitude research has not seen much methodological innovation since the 1960s. Social psychology, by contrast, has witnessed an explosion of new attitude measures in recent years. This section will start out by situating this methodological stagnation in linguistic attitude research and give a compact overview of the methods this field has traditionally relied upon (Section 2.1).

Following this overview, we will turn to the new dynamics that have recently begun to break the methodological inertia in the field.

Section 2.2 elaborates on the new implicit attitude measures developed in social psychology by describing some of their key characteristics. In that context, we will elaborate on the concept of implicitness and the opportunities that engaging with this psychological concept may hold for linguists (Section 2.2.1). The rest of the section is dedicated to the three measures directly relevant to the studies presented in the following chapters: affective priming (AP, Section 2.2.2), the Implicit Association Test (IAT, with particular attention to the personalized version of the test, Section 2.2.3) and the Relational Responding Task (RRT, Section 2.2.4). The chapter concludes with a comparative review of the three methods which focuses particularly on aspects that may have consequences for their implementation in sociolinguistic attitude research (Section 2.2.5).

2.1 Attitude measurement in linguistics

Language attitude research has been making use of the same methods, virtually since the 1960s[4] when the field's best known method was introduced: the matched guise technique (MGT, Lambert et al. 1960)[5]. Apart from the MGT, linguists have mostly relied on questionnaires for attitude measurement. That linguistics has not ventured much further than these two methods (until quite recently) becomes clear when we consult a number of recent reference works on sociolinguistics and language attitude research in particular. In the 2010 edition of Peter Garrett's reference book on language attitudes, MGTs and questionnaires form the main representatives of their respective type of method: indirect vs. direct. Garrett adds a third type of methods which he refers to as societal treatment. Techniques belonging to the latter rely on the analysis

[4] A notable exception is the introduction of folk linguistics in the 1980s which brought with it a number of new methodological approaches (e.g. Preston 1982; ibid. 1989; ibid. 1993; ibid. 1999; see below).

[5] Note that the matched guise technique was actually introduced by social psychologists, rather than linguists.

of different sorts of (public) documents to investigate how languages are viewed in society. When we look at recent handbooks on sociolinguistics, we observe that they refer to (a subset of) the same traditional attitude measures discussed by Garrett (2010) with no mention of any recent developments or additions of new approaches (Meyerhoff 2006; Garrett 2007; Wolfram 2011[6]). This suggests that methodological innovation has been limited for language attitude research in the past few decades. Yet that state of methodological inertia seems to be coming to an end. If we continue using reference works as an indicator of general trends in the field, this new methodological dynamic is exemplified in a recent edited volume by Prikhodkine and Preston (2015). Although methodology is not the main focus of the publication, it showcases two trends: innovations in traditional methods (as already indicated in Wolfram 2011) and novel methods inspired by social psychological attitude measures. In what follows, we will briefly discuss the traditional methods used in linguistic attitude research following Garrett's three-way distinction (indirect, direct and societal treatment) and give examples of the most commonly used techniques within each category. Following this discussion, we will take a closer look at the recent innovations in language attitude research.

The first and perhaps most prototypical type of traditional techniques are the indirect methods where researchers measure language attitudes without directly asking participants to report them (Garrett 2010). Garrett (2010) equates indirect approaches with the *speaker evaluation paradigm* which encompasses matched and verbal guise techniques (MGT and VGT henceforth). In these experimental methods, participants are asked to listen to a number of audio samples that only differ in the languages or language varieties they contain. Listeners then rate the speaker(s) of these samples on a number of personality traits. These ratings are taken to reflect their language

---

[6] Although Wolfram (2011: 307) indicates that '[t]he development of experimental methods in perceptual studies is one of the fastest growing areas of sociolinguistics […]', he seems to be referring mainly to technical innovations (concerning for instance speech synthesis) within the more traditional methods, and to more rigorous experimental designs and application of statistical methods again in relation to the same traditional methods, rather than to the introduction of novel measurement techniques.

attitudes. Crucially, respondents rate the speaker, not their speech. The aim of asking about the speaker rather than their speech is to avoid participants becoming aware of the fact that the researcher is measuring their language attitudes. The difference between the MGT and VGT is that the former relies on a single speaker to produce different languages or language varieties, while the latter uses multiple speakers. The motivation for using a single speaker for multiple language varieties in the MGT is to keep voice quality constant, as differences in voice quality can influence the ratings of the speech samples and thus create a confound. However, it may be difficult to find a speaker who is equally fluent in all languages or varieties included in a study. This often leads to contrived or unauthentic renderings of one or more of the languages or varieties. To deal with this problem, researchers may instead opt for the VGT where multiple speakers are used. The choice between the MGT and VGT thus comes down to weighing up control and authenticity.

Although the MGT and VGT have been very productive and have proven their usefulness for linguistic attitude research, they come with a number of problems. As illustrated by the choice between the MGT and VGT, many of these issues relate to finding a balance between controlling for variables that may influence the outcome of the experiment and making sure participants are presented with more or less natural sounding fragments where the attention is not drawn to the language. One example of such a problem is the content of the speech samples that are used in a MGT or VGT experiment. The topic needs to be as neutral[7] as possible and variation between the different guises should be limited. One solution is to have the speaker(s) read an identical text in each of the guises, which ensures maximal control over the content of the audio samples. However, hearing the same text over and over again with the language as the only difference, will most likely draw attention to the fact that the researcher is interested in the participant's language attitudes. This is of course the exact opposite of what the speaker evaluation paradigm aims to achieve. Moreover, while reading a text may assist in limiting the variation in content,

[7] Producing a neutral text is in itself a challenge, which may even be unachievable (Garrett 2010: 59).

it can also lead to a more contrived way of speaking, which again could have a reflection in respondents' ratings.

Another issue related to both the MGT and VGT is that they traditionally rely on a number of semantic differential scales including traits like 'intelligent – unintelligent' or 'kind – unkind' to measure attitudes. This means that the researcher has to decide in advance which traits to include in the study. By doing so, one may miss out on aspects of the social meaning of a language (variety) that were not known or suspected to be relevant beforehand. Hence, thorough exploratory research is required before conducting MGT or VGT experiments of this type. An overview of further criticisms associated with the MGT and VGT can be consulted in Garrett (2010: 57-59) or Bradac et al. (2001: 139-141).

The second of Garrett's three types of language attitude measures comprises a variety of techniques that directly asks respondents about their attitudes towards a linguistic phenomenon. Questionnaires were already mentioned above, but other examples of direct measures include interviews and methods emanating from Dennis Preston's folk linguistics (e.g. Preston 1989)[8]. The best-known technique from the latter paradigm is probably the 'draw a map' task. For this task, participants receive a map with a minimum amount of landmarks. They are then asked to draw a circle around areas where people speak in a similar way. They are also invited to provide labels for these ways of speaking, as well as other qualifications or characteristics they can come up with. This type of task represents a more exploratory approach than for instance the MGT, which makes it an interesting starting point for a large-scale, multimethod study. It allows one, for example, to 'harvest' adjectives that can be used in semantic differential scales in subsequent questionnaires or MGT experiments. Another advantage of this technique is that it offers an insight into the non-linguist's cognitive reality of (regional) varieties: laypeople may distinguish varieties where dialectologists/ sociolinguists do not and vice versa. Furthermore, the 'draw a map' task can be highly useful to find out how

[8] Note that folk linguistics is preceded by an older tradition of perceptual dialectology (e.g. Weijnen 1946).

non-specialists label varieties. As will be discussed in Section 2.2.5 and again in Chapter 4, the choice of a label to describe a language (variety) is not a trivial decision when designing a language attitude experiment. In addition, the 'draw a map' task may give some indications as to which features of a variety are salient for linguistic laymen. A potential drawback is that the method strongly focuses on geographic space, which makes it less attractive for studies without a regional component.

Another example of a technique from the direct methods category is the keyword task (e.g. Garrett et al. 2003; Garrett et al. 2005). When this method is applied, participants are asked to jot down a number of associations in response to a language (variety) which can be presented to them in the form of an audio sample or a label. They have to do this as quickly and as spontaneously as possible. Like with the 'draw a map' task, the keyword method can be an interesting approach to explore attitudes and tap into potential dimensions that structure them. In this way, it can provide input for other methods like the MGT. A possible drawback of the keyword approach is the need for post-hoc interpretation of the keywords by the researcher. Participants usually write the keywords down without adding any context or explanation. Hence, some caution is warranted when analysing them as respondents may have had various interpretations in mind which the researcher can never retrieve (Garrett et al. 2005). For example, if a variety is qualified as 'posh', this may be interpreted as an indication of prestige. However, Garrett at el. (2005) point out that it may just as well have been intended as a negative qualification signalling arrogance or snobbism. Without further context, it is impossible for the researcher to be certain what was meant by the participant.

Garrett's (2010) final type of methods is referred to as 'societal treatment' and contains a heterogeneous set of approaches to language attitudes. Studies that fall into this category use various types of public sources to investigate how languages or varieties are treated in different communities, or rely on ethnographic techniques (Garrett 2007). For example, some studies analyse policy documents (e.g. Delarue 2011), while others use advertisements (e.g. Bishop et al. 2005b) to learn more about language attitudes and ideologies in society.

Notwithstanding a number of smaller adaptations and additions to the core methods discussed above, the field of language attitude research has only recently witnessed more profound methodological innovations (Preston 2016). On the one hand, linguists are actively improving and refining some of the traditional methods, and on the other, new methodologies are being introduced inspired by discourse-based research in linguistics or spurred on by advances in social psychological attitude research.

There are various examples of studies providing new impulses for traditional language attitude research. Grondelaers & Speelman (2015) and De Pascale (forthcoming), for instance, apply new statistical techniques to analyse responses from keyword tasks. Staum Casasanto and collaborators (2015) use images in a forced choice task to replace conventional Likert scale ratings with verbal descriptions of social traits in speaker evaluation experiments. Llamas & Watt (2014) present innovative direct attitude research based on Visual Analog Scales rather than traditional Likert scales. And Montgomery & Stoeckle (2013) introduce innovations in processing data collected through 'draw a map' tasks.

The innovations in language attitude research are not restricted to improving existing methods, an important part of which are quantitative in nature. In recent years, the study of language attitudes is also witnessing a reinforcement of qualitative techniques. These find their roots in discourse-based linguistic traditions (e.g. Liebscher & Dailey-O'Cain 2009).

In addition to new approaches to older methods, and to extending the array of qualitative techniques, linguists are gaining interest in exploring measures recently developed in social psychology (see Section 2.2 for a detailed introduction to these methods). Only a few of those social psychological measures have been explored in linguistic attitude research so far, the most popular being the Implicit Association Test (IAT, Greenwald et al. 1998; see Teige-Mocigemba et al. 2010 for a more recent introduction). The IAT is a reaction time based categorisation task that measures the association between two binary concepts (e.g. candy/vegetables and good/bad) (see Section 2.2.3). So far linguists have employed the IAT to study the following aspects of language varieties and linguistic variants: their evaluation (Babel 2010; Chu

2013; Redinger 2010; Pantos & Perkins 2012; Lee 2015; Watt & Llamas 2015; Loudermilk 2015; Leinonen 2016; McKenzie 2017), their social meaning and indexicality (Campbell-Kibler 2012; ibid. 2013; Llamas et al. 2016; Hilton et al. 2016), and their salience (Leinonen 2016). To the best of our knowledge, the only other implicit reaction time based method from social psychology that linguists have explored is Affective Priming (AP, Fazio et al. 1986; Speelman et al. 2013). In the next section, a brief introduction to the recent methodological innovations in social psychology is given.

2.2 Attitude measurement in social psychology

In sharp contrast with the relative stagnation in language attitude research, social psychology has witnessed a substantial increase in the development of new attitude measures in the past two decades. As a field, social psychology takes a strong interest in the study of attitudes and has developed a vast range of measurement techniques to pursue that study. These methods are highly diverse and range from traditional questionnaires to reaction time based tasks and the use of brain imaging techniques like Event Related Potentials (ERP). Many ways of classifying this wide variety of measures have been proposed and it would go beyond the scope of this thesis to fully present even some of them[9]. However, many of the recently introduced social psychological attitude measures have been classified as implicit. Hence, in what follows, we will discuss the distinction between implicit and explicit attitude measurements from a social psychological perspective. Our focus on implicitness is not only motivated by its relevance to the methods introduced in Sections 2.2.2-2.2.4, we are also convinced that the social psychological approach to thinking about this concept may have some value for sociolinguists studying language attitudes. This will be elaborated upon in Section 2.2.1.

An attitude measure is defined as implicit if the to-be-measured attitude 'influences participants' responses on the task in an automatic fashion' (Gawronski & De Houwer 2014, 284). In other words, implicitness is defined

[9] For an excellent exercise in classifying attitude measures with a focus on implicit measures, we refer the reader to De Houwer & Moors (2010).

in terms of automaticity (De Houwer & Moors 2007). The concept of automaticity can be broken down into a set of characteristics (Bargh 1994; Moors & De Houwer 2006). A process is called automatic if it occurs under circumstances where participants have no control over the process, little time, few cognitive resources, no particular intentions and limited awareness (De Houwer et al. 2009; see also Moors & De Houwer 2006 for a critical review of the concept of automaticity and its different features). These characteristics are gradual in nature (Moors & De Houwer 2006) and do not necessarily co-occur (De Houwer et al. 2009). This means that various measures can be implicit in different ways and to a different extent. Processes that do not possess any of these automaticity features can be classified as explicit.

De Houwer and colleagues (2009) emphasize that the term 'implicit' is best reserved to talk about the processes that lead to a certain response in a certain task. If one wants to refer to the measurement procedure that is used to get a glimpse of these attitudes, the terminological pair 'direct-indirect' is more appropriate[10]. A measure is qualified as direct if respondents self-assess their attitudes and if they are directly asked about the attitudes (De Houwer & Moors 2010). An indirect measure, by contrast, can or cannot be self-assessed, but does not directly question respondents about the attitudes under study. By way of illustrating these terms, let us take a look at the MGT (cf. Section 2.1). Although the MGT requires participants to engage in self-assessment, it does not ask them directly about language. Hence, the MGT fulfils the criteria to be classified as an indirect measure. To be able to say that the MGT relies on implicit processes, one would have to investigate which automaticity features apply to what extent, which – to the best of our knowledge – has not been attempted yet.

The advantage of indirect measures is that they may offer a solution when direct self-report measures are insufficient, for instance when attitudes

[10] Despite this plea for clear terminological usage by De Houwer et al. (2009), the terms 'implicit' and 'indirect' are often used indiscriminately in social psychological attitude research. In this thesis we maintain the distinction as much as possible, but if the use of the other term is more common in the literature, we will opt for that term to avoid confusion and maintain the connection with previous work (cf. Rosseel et al. 2014).

are investigated that participants are unwilling or unable to comment on (Gawronski & De Houwer 2014). In certain attitude domains, expressing particular views may be perceived by respondents as socially undesirable or, in other cases, they may simply not be aware of the attitudes they hold. In circumstances like these, alternatives to direct self-report measures may prove highly useful.

We indicated at the beginning of the section that a considerable number of measures have been developed over the past 20 years that have been presented as implicit. To name just a few, there is the Go/No-go Association Task (GNAT, Nosek & Banaji 2001), the Affect Misattribution Procedure (AMP, Payne et al. 2005), the Implicit Relational Assessment Procedure (IRAP, Barnes-Holmes et al. 2010), the Approach-Avoidance Task (Chen & Bargh 1999) and the list goes on. For the studies presented in Chapters 2-4, three of these implicit measures are of particular relevance: Affective Priming (AP, Fazio et al. 1986), the Implicit Association Test (IAT, Greenwald et al. 1998 – especially in its personalized variant, the P-IAT, Olson & Fazio 2004) and the Relational Responding Task (RRT, De Houwer et al. 2015). The P-IAT is explored in the studies reported in Chapters 2 and 3, while the RRT is the implicit measure employed in Chapter 4. Although AP is not used in any of these studies, it is crucial for the one presented in Chapter 2. That study reports a P-IAT experiment that was modelled on Speelman et al. (2013), the first linguistic study to employ AP to measure language attitudes. Chapter 2 compares the outcome of the two implicit measures alongside an explicit measure.

Before we continue with a discussion of the specifics of each of these three methods, it may be useful to briefly indicate what they have in common. AP, the (P-)IAT and the RRT are all reaction time based categorisation tasks. Participants are presented with stimuli they have to categorise according to the task they are set. What is measured during these categorisation tasks is whether participants categorise stimuli correctly and how quickly they do it. Very broadly speaking, the idea behind all three of the methods is the following: if there is compatibility between the participants' attitudes and some aspect of the categorisation task, they will be able to respond faster than when their

attitudes are incompatible with that aspect of the task. How this works precisely for each of the three methods will be explained in Sections 2.2.2-2.2.4.

Besides the fact that AP, the (P-)IAT and the RRT involve categorising stimuli and measure reaction times, they differ substantially from one another. These differences are situated on various levels ranging from how the measures are implicit, to which processes underlie them as well as their procedural aspects. In the remainder of this chapter, we will take a procedural point of view to first discuss each of the methods individually (Sections 2.2.2, 2.2.3 and 2.2.4) and then evaluate the three methods on a number of criteria in a comparative review (Section 2.2.5). Based on this review, we will attempt to come to a preliminary conclusion about the potential of these social psychological measures for sociolinguistic attitude research. We focus on the procedural aspects of the three methods under study, because our first goal is to introduce these psychological methods to sociolinguists and to offer information that may help them decide whether and how these measures can be of use for their linguistic research. For a thorough analysis of the implicitness of several social psychological attitude measures including AP and the IAT we refer to De Houwer et al. (2009), for the RRT we recommend De Houwer et al. (2015) and Heider et al. (2016). For a discussion of the processes underlying the different methods, see Gawronski et al. (2011) for the IAT and Spruyt et al. (2011) for AP. Given the novelty of the RRT, the documentation of its underlying processes is still rather limited.

2.2.1 Implicitness as an opportunity for sociolinguistics

Before moving on to the detailed discussion of our three social psychological implicit attitude measures, let us briefly go back to linguistic attitude research and our choice to adopt a more psychological approach to implicitness. As became clear in Section 1, linguists appear to make a similar distinction between implicitness and explicitness (cf. overt/covert, indirect/direct, deep/shallow evaluations), but a clear definition or terminological unity

comparable to that in social psychology[11] seems to be lacking as of yet. Given the absence of clear definitions for the linguistic concepts, it is hard to judge to what extent they overlap with the social psychological interpretation presented above. As stated in Section 1, the focus in linguistics seems to be on awareness, rather than framing implicitness as a multidimensional and gradual concept like in social psychology. In the studies reported in this thesis, we will mainly work with the social psychological interpretation of implicitness, given that we work with methods from that field which sparked much of the theorising around implicitness. We do not argue, though, that linguists should adopt this view of implicitness unquestioningly. Rather, what we would like to propose is that linguists may use the psychological conceptualisation as starting point or inspiration to further develop the concept, building on our own sociolinguistic theories. Besides conceptual clarity, there are two additional reasons why we believe linguists should further develop their thinking about implicitness. Firstly, clear conceptualisation of implicitness can help to determine its theoretical importance for language variation and change. Secondly, it may lead to a better understanding and use of the different methods at the sociolinguist's disposal for measuring language attitudes.

When it comes to the theoretical importance of implicitness, it has been claimed that implicit evaluation can provide important insights to understand language variation and change (e.g. Kristiansen 2010b; Preston 2013). However, studies like Soukup (2013b), which showcases the use of an open guise technique (where participants are aware of the fact that one speaker uses different language varieties), claim to be able to explain language variation in certain contexts. This may raise questions like: do we always need implicit measures? What is the theoretical significance of implicitness in the study of language variation and change? Should it occupy a privileged position when it comes to explaining the driving force behind language change as suggested by

[11] Note that social psychologists are not always consequent in their terminology concerning implicitness either (Gawronski & De Houwer 2014), nor do they all adhere to the same definitions and theories when it comes to implicitness (Moors & De Houwer 2006). Yet considerable efforts are made to communicate transparently about the concepts under study and attempts are made to arrive at precise definitions and theories that can be compared and empirically tested.

Kristiansen (2010b) contrary to, for instance, Labov's (2001) current more anti-subjective position?

In addition to theoretical importance, challenging the linguistic conception of implicitness has important methodological consequences. If we ask ourselves the question what exactly we mean by implicitness, and if we should find that it is a multifaceted concept, we should also ask ourselves *which* aspects of implicitness we are measuring using certain methods and tools. This goes for methods traditionally used in sociolinguistic research, but the question is especially relevant in the context of the recent interest in social psychological measures to study implicit associations (e.g. Campbell-Kibler 2012). When we adopt new methods it is crucial to question what it is exactly that these tools measure, how these methods fit in with sociolinguistic conceptions of attitudes and social meaning, and how the measurements compare to the ones obtained from more traditional sociolinguistic tools (e.g. the MGT or VGT). As linguists we should take on the challenge to critically reflect on what we do with these measures, and how it fits in with our traditional methods and with sociolinguistic interpretations of concepts like 'attitudes' and 'indexicality'. As indicated in Section 1, in a number of more recent studies, some of which employ social psychological implicit measures, a trend towards more critical reflection on implicitness is already observable (e.g. Campbell-Kibler 2012; Preston 2015).

2.2.2 Affective priming

The affective priming task (Fazio et al. 1986) involves the sequential presentation of two stimuli on a monitor (Spruyt et al. 2011). The first one is referred to as the prime, the second one is the target stimulus. Participants are instructed not to react to the prime which is only briefly shown. Sometimes the prime is even presented subliminally (Gawronski & De Houwer 2014). The target stimulus which follows the prime is to be categorised as quickly as possible as being either positive or negative, by pressing one of two response

keys[12]. One such sequence of a prime and target stimulus is called a trial. The idea is that when prime and target are related in terms of valence, categorisation will be faster than when they are unrelated in that respect. In other words, participants will categorise a target stimulus faster when it is preceded by a congruent prime (positive prime + positive target or negative prime + negative target) than when it is preceded by an incongruent prime (negative prime + positive target or positive prime + negative target). This is summarised in Figure 1 where the images represent the three stages of a trial: the prime presentation, followed by the target presentation and finally the respondent's categorisation of the target stimulus. The table below the images lists the potential combinations of primes and targets and their impact on the response time.

prime	target	congruence	response speed
+	+	congruent	faster
+	-	incongruent	slower
-	+	incongruent	slower
-	-	congruent	faster

Figure 1 - Schematic representation of the Affective Priming procedure

To illustrate this, say we are interested in measuring whether people hold positive or negative attitudes towards certain types of food (e.g. vegetables vs. chocolate). In that case, we could use pictures of vegetables and chocolate as primes. Pictures or words for which norm data has been collected can

[12] Some variations on AP include alternative categorisation tasks, like naming positively or negatively valenced pictures (Spruyt et al. 2007). For a detailed overview of all potential procedural variants of the task, see Wentura & Degner (2010).

function as targets, so we can be certain that these pictures/words are universally perceived as positive or negative. We then analyse respondents' reaction times to see which combinations of foods and target pictures/words lead to faster or slower reaction times. It is advisable to include a neutral type of primes in the experiment which can be used as a baseline when analysing reaction times (Wentura & Degner 2010).

In addition to relations of valence between primes and targets, the priming paradigm also allows the study of other relations between the two stimuli (Gawronski & De Houwer 2014). In that case, we are not dealing with affective priming anymore, but with a different technique called semantic priming (e.g. Banaji & Hardin 1996; Wittenbrink et al. 1997). Semantic priming could for example be used to study stereotypical traits of certain social groups or the association of particular professions with gender. Taken together, these different types of priming techniques are referred to as the sequential priming paradigm (Spruyt et al. 2011).

Besides its flexibility regarding the relationship between prime and target categories, AP is also flexible concerning the type of stimuli that are used, as long as they can be presented briefly (Spruyt et al. 2011). In the vegetables-chocolate example above, we already referred to pictures and words, but auditory stimuli can be used as well, which offers interesting opportunities for linguists. The use of auditory primes was explored by Speelman et al. (2013) who introduced AP to the field of language attitude research.

An aspect for which AP is sometimes criticized is its reliability which is often below par. This issue, as well as other aspects of AP that may make it a more or less desirable option for linguists to use as an attitude measure will be discussed in detail below (see Section 2.2.5).

2.2.3 Implicit Association Test

The IAT (Greenwald et al. 1998) measures the association between a binary target concept (e.g. language variety: variety A vs. variety B) and a binary attribute concept (e.g. valence: good vs. bad) by comparing reaction times in a number of computer based categorisation tasks. For each of these four concept categories, a number of stimuli are required that are representative of their

respective category. During the IAT, participants are asked to categorise the target and attribute stimuli according to the corresponding target and attribute categories respectively. This is done by pressing one of two response keys representing the categories involved in the experiment. The mapping of the categories onto the response keys is indicated with labels in the top corners of the computer screen throughout the experiment, so participants do not need to memorize the mappings (see Figure 2). When participants press the wrong key, they receive error feedback, for instance in the form of a red cross that appears in the middle of the screen until they correct themselves by assigning the stimulus to the right category.

An IAT is made up of a series of trials which each require the categorisation of one stimulus. These trials are divided into seven blocks. The first two blocks of trials are practice blocks which aim to familiarize the participant with the stimuli used in the experiment, the categorisation task and the mappings of the response keys. The first block consists of target stimulus discrimination: in each trial participants indicate which of the two target categories a stimulus belongs to (see block 1 in Figure 2). The second practice block involves the categorisation of attribute stimuli according to the attribute categories 'good' vs. 'bad' (see block 2 in Figure 2). The third and fourth block are two identical experimental blocks. They combine the target and attribute discrimination practiced in the first two blocks. Both target and attribute stimuli have to be categorised in these blocks using the two response keys on which both a target and an attribute category are mapped now (e.g. 'variety A' + 'bad' for the left-hand key, and 'variety B' + 'good' for the right-hand key, see blocks 3 and 4 in Figure 2). Note that each stimulus belongs to one of the four categories only. This set of experimental blocks is followed by another practice block requiring target discrimination. This fifth block is identical to the first block except that the category labels mapped onto the response keys have swapped sides (see block 5 in Figure 2). If, for example, the left key corresponded to variety A in the first block, it will now correspond to variety B and vice versa for the right key. Note that this block usually contains twice as many trials as the first practice block. This gives participants ample time to get used to the new configuration which should help to avoid compatibility order

effects[13] (Teige-Mocigemba et al. 2010; Gawronski et al. 2011). The final two blocks are again identical experimental blocks and contain trials in which either target or attribute stimuli are to be categorised. For the target categories, the response key mappings from the fifth block are retained, while the mappings for the attribute categories are kept constant throughout the experiment. This results in a response key mapping in blocks 6 and 7 that is the reverse of the mapping in blocks 3 and 4, the other set of experimental blocks (see blocks 6 and 7 in Figure 2).

This inverse response key mapping in the two sets of experimental blocks is primordial to the mechanism behind the IAT. Categorisation of the stimuli is assumed to be easier if the responses mapped onto the same key are congruent according to one's attitudes. Conversely, when the mapping of categories onto the response keys is incongruent with one's attitudes, categorisation of the stimuli will be harder. For instance, for a participant with positive associations with variety A, but negative associations with variety B, stimulus discrimination will be easier if target stimuli representing variety A and intrinsically positive attribute stimuli, like pictures of a smiling child or a sunny beach, are categorised with the same key. Yet, it will be more difficult for that person if 'variety A' and 'bad' are assigned to the same response key. Easier categorisation will lead to faster reaction times, while a harder categorisation task will slow down responses. By comparing reaction times between the two sets of experimental blocks, we can determine which categories participants associate more strongly.

---

[13] The IAT has been shown to suffer from block order effects: if the congruent block (i.e. where the key mappings correspond to the participant's attitudes) precedes the incongruent block (i.e. where they do not match), the IAT effect tends to be larger. This issue has been tackled in several ways. For one, IAT variants have been developed that deal with compatibility order effects by getting rid of the block structure (e.g. the Single Block IAT, SB-IAT, Teige-Mocigemba et al. 2008; the Recoding Free IAT, IAT-RF, Rothermund et al. 2009). Another way of diminishing block order effects is to increase the number of trials in the fifth block of the IAT from the traditional 20 trials to 40 (Nosek et al. 2005). In the studies presented in Chapters 2 and 3, we have opted for this last solution, given that the SB-IAT and IAT-RF tend to be even more cognitively demanding for participants than a standard or personalized IAT.

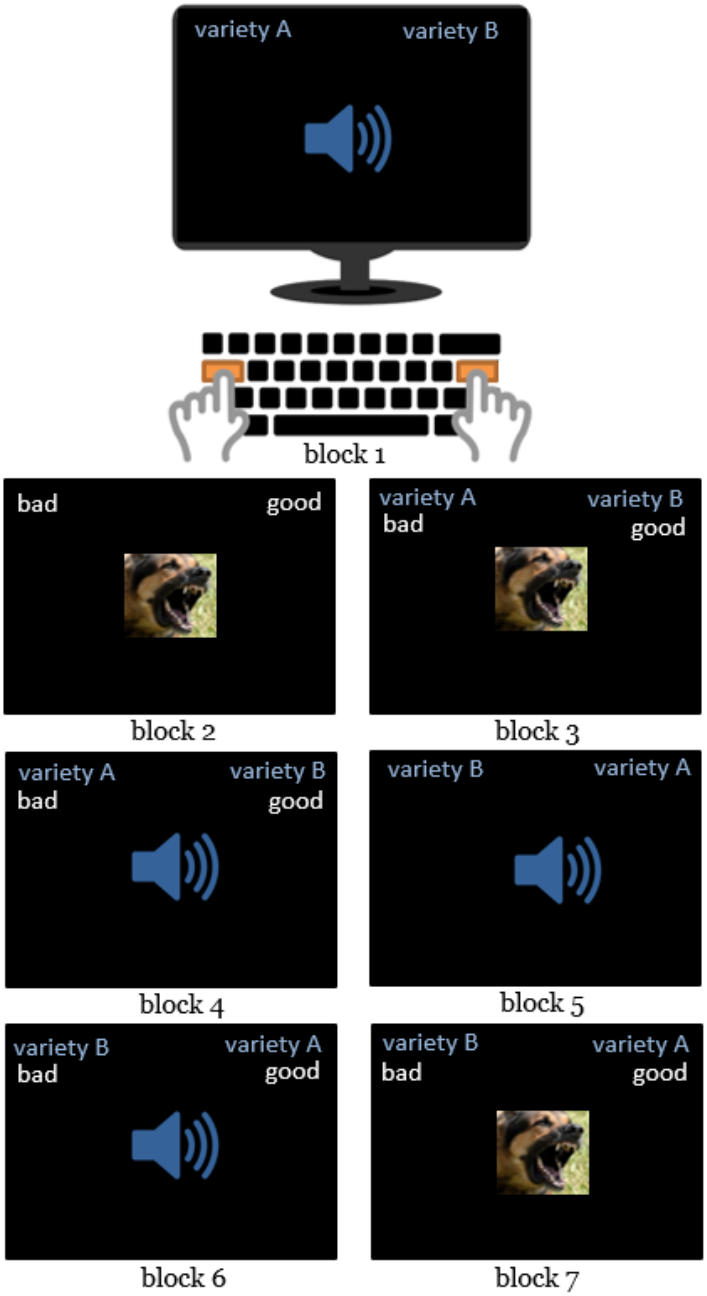


Figure 2 - Screenshots with an example of a trial from each block of an IAT. The example for block 1 also illustrates the experimental setup as used in Chapters 2 and 3.

Just like AP, the IAT is used very frequently in social psychology and has been successfully applied to study a wide variety of topics, ranging from racial stereotypes (e.g. Greenwald et al. 1998) to addictive behaviour (e.g. Houben & Wiers 2006) and advertising (e.g. Maison et al. 2004). One reason for the IAT's

popularity in social psychology, is that it has been shown to have good psychometric qualities (Nosek et al. 2007). In addition, the measure is quite flexible, for instance in the type of stimuli it allows (written words, images, sound clips, etc.) and the type of associations that can be measured (i.e. not restricted to good/bad associations). However, the method comes with a number of characteristics which are sometimes considered limitations. One such characteristic is the fact that the concepts studied in the IAT have to be presented as binary categories. To make up for this potential limitation, a number of variants of the traditional IAT have been developed. Measures like the Single Target IAT (Wigboldus et al. 2004) and the Single Attribute IAT (Penke et al. 2006) allow for non-binary concepts.

Another problem that has been linked to the IAT is the measurement of extra-personal associations instead of personal associations (although this distinction is not uncontroversial, Gawronski et al. 2008; De Houwer et al. 2009). Personal associations refer to preferences endorsed by an individual. Extra-personal associations, on the other hand, are societal views which are present in memory, because they are frequently encountered, but they are not necessarily endorsed by the individual. For example, for someone who dislikes vegetables, a traditional IAT may still return positive associations with that type of food, because that person will have been repeatedly confronted with the information that vegetables are healthy and good for you, for instance in school or through government campaigns. To deal with this potential disadvantage, the Personalized IAT (P-IAT) was developed (Olson & Fazio 2004). In the studies presented in Chapters 2 and 3, the personalized variant of the IAT will be applied for the first time – to the best of our knowledge – in the context of linguistic attitude research. The features that personalize an IAT are, firstly, the use of attribute labels that refer directly to the participant's opinions, such as 'I like' and 'I don't like' instead of 'good' and 'bad', and secondly, the omission of corrective error feedback after participants press the wrong key.

Before further considering potential advantages and disadvantages of the IAT paradigm in relation to language attitude research in Section 2.2.5., we will now first turn to the RRT.

2.2.4 Relational Responding Task

The RRT is a method that was introduced only very recently (De Houwer et al. 2015), but is already beginning to be applied in various domains of psychology (e.g. parenting cognition and alcohol use in Koning et al. 2016, or sexual self-concept in Dewitte et al. 2017). Although the structure of a RRT with its seven blocks is reminiscent of the IAT, a crucial difference between the IAT (/AP), and the RRT is that the former aim to measure associative relations, while the latter focuses on propositional ones (De Houwer et al. 2015). Traditionally, many psychologists adhere to a view of memory as consisting of an associative component and a propositional one, with the former being prone to automatic processes, but not the latter, which is considered to operate solely under controlled circumstances (Hughes et al. 2016, 634). De Houwer et al. (2015), however, argue that propositional relations may be activated automatically as well, and developed the RRT to measure these relations. Psychologists refer to these propositional relations as beliefs (De Houwer et al. 2015). To make this more concrete, consider the example where a researcher is interested in studying body image. The IAT could be used in this case to measure the association between the concepts 'self' (I vs. other) and 'weight' (fat vs. thin). If a participant's results show an association between the self and thin, the researcher lacks information on how the two concepts are related for this participant. That is to say, in research on body image, the difference between 'I am thin' and 'I want to be thin' is crucial and the IAT cannot distinguish between the two[14]. This is where the RRT can make a difference. The RRT does not measure the association between two loose concepts (like 'I' and 'thin'), rather, it focuses on propositions (like 'I am thin' or 'I want to be thin'). The idea is that participants categorise a number of statements as being true or

[14] Note that a P-IAT comes closer to making such distinctions as it uses category labels that include a more precise proposition (e.g. 'I like'/'I don't like'). Although the P-IAT is designed to measure attitudes (i.e. associations between the attitude object and the attribute concept), the use of 'I like' as an attribute label renders it more propositional in nature and makes it lean towards a belief. This illustrates that the distinction between beliefs and attitudes can be fuzzy (Albarracín et al. 2005). Hence, we will not adhere to a strict distinction between the two.

false, not based on what they think themselves, but reacting *as if* they adhered to a certain belief (which is imposed on them by the researcher). If this imposed belief matches their own belief, they will be able to categorise the statements faster than when this is not the case.

As mentioned above, the RRT has a seven block structure: two practice blocks followed by two experimental blocks, another practice block and finally, two experimental blocks again (see Figure 3). The first practice block presents the participants with a series of inducer trials. The stimuli in these trials are usually synonyms of true and false (e.g. correct, right, incorrect, wrong) and have to be categorised as such using two response keys. The meaning of the response keys ('true' and 'false') is indicated in the top corners of the screen and remains the same throughout the RRT. The main function of the inducer trials is to remind participants of the meaning of the response keys and to avoid response recoding, i.e. categorising based on another feature of the stimuli than whether they are synonyms of true or false (De Houwer et al. 2015; Dewitte et al. 2017).

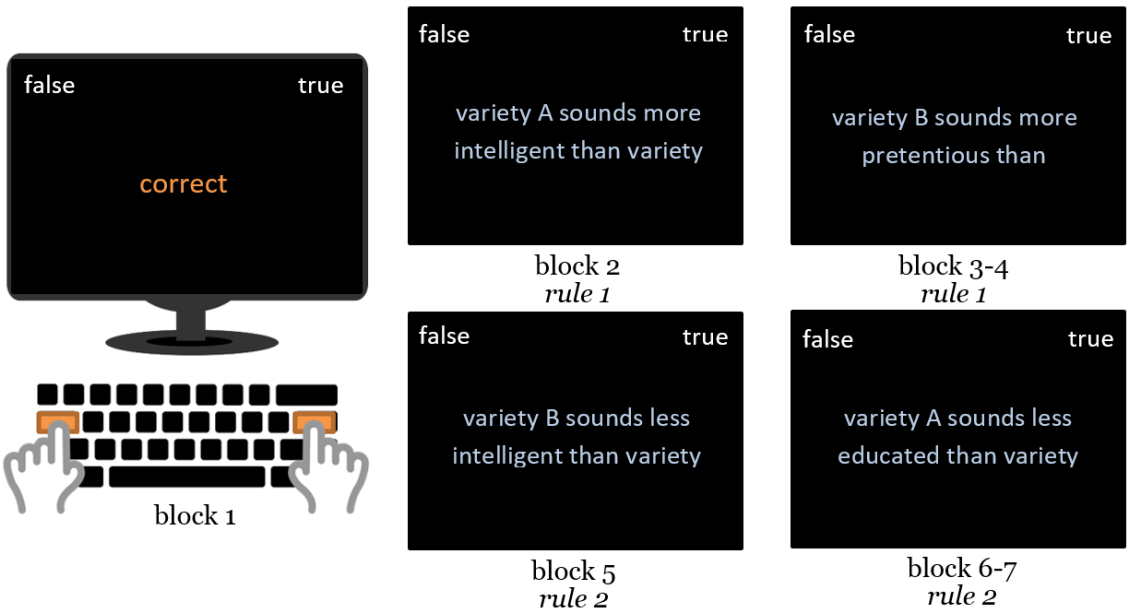


Figure 3 - Screenshots with an example of a trial from each block of a RRT. The example for block 1 also illustrates the experimental setup as used in Chapter 4.

The second block consists of a series of trials that each present a statement containing a proposition about the attitude object. For instance, if one wants to study prestige evaluations of language varieties A and B, statements could include 'variety A sounds more intelligent than variety B' or 'variety B sounds more educated than variety A'. These statements are to be categorised as true or false as quickly as possible, using the same two response keys as in block 1. To decide whether a proposition is true or not, participants are instructed not to rely on their own beliefs, but to respond *as if* they held the belief formulated in rule 1. So say rule 1 in our example is 'variety A sounds more prestigious than variety B', then participants would have to categorise 'variety A sounds more intelligent than variety B' as true and 'variety B sounds more educated than variety A' as false. This block allows participants to practice responding according to rule 1.

In blocks 3 and 4, the first pair of experimental blocks, inducer trials and statements are randomly intermixed. Respondents still have to indicate for the inducer trials whether they are a synonym of true or false. For the statements, they keep following rule 1 to decide whether a statement is correct or not.

Block 5 constitutes another practice block containing the same statements as in blocks 2-4, but no inducer trials. Now participants are instructed to respond to the propositions as if they believed rule 2. In our example about language varieties A and B, rule 2 would be 'variety B sounds more prestigious than variety A'. So now the correct response to 'variety A sounds more intelligent than variety B' would be false and to 'variety B sounds more educated than variety A' it would be true.

After practising rule 2 in the fifth block, block 6 and 7 are again experimental blocks containing both statements and inducer trials. Now participants keep on following rule 2 for the categorisation of the statements. As indicated above, the idea is that respondents will be able to react faster to the statements when they are allowed to follow the rule that expresses a belief congruent with their own. So the reaction times from the experimental blocks are compared to see whether a participant's beliefs align with rule 1 or rule 2. In our example, that means that a participant who believes variety A to be more prestigious reacts faster in blocks 3 and 4, because their belief matches rule 1.

For a participant who thinks variety B sounds more prestigious, response latencies will be shorter in blocks 6 and 7 given that rule 2 is to be followed.

Social psychologists also have another implicit measure at their disposal that uses propositions and hence solves the problem of unspecified relations between concepts: the Implicit Relational Assessment Procedure (IRAP, Barnes-Holmes et al. 2010). Notwithstanding multiple successful applications, the IRAP's procedure is highly complex. As a result, some participants simply fail to complete the experiment correctly (De Houwer et al. 2015). Comparatively, the RRT – although, admittedly, itself rather complicated for participants to understand – offers a simpler procedure and may in that respect be more useful with certain groups of participants than the IRAP.

2.2.5 Comparative review: AP, the IAT and the RRT

After introducing AP, the IAT and the RRT, we now turn to a comparative review of the three methods from a procedural perspective. Although we have already briefly touched on a number of aspects that may make these methods attractive or challenging to work with, in this section we will systematically review and compare a number of their procedural characteristics. In order to do so, we have composed a set of 7 criteria (see Table 1) on which to evaluate and compare the three methods. This review is by no means exhaustive, but selects a number of aspects of the methods we deemed to be worth considering for sociolinguists who are interested in using one of the three measures. After reviewing the methods on all criteria, we will attempt to formulate a preliminary evaluation of the potential of these social psychological implicit attitude measures for linguistic research into the social meaning of language variation.

1.A STIMULUS MODALITY & 1.B LENGTH | The first criterion to be discussed is **flexibility** and encompasses a varied range of aspects (1.A - 1.H). Firstly, let us discuss whether there are any restrictions on the stimuli that can be used in experiments based on AP, the IAT or the RRT. The former two measures have been shown to allow virtually any type of stimulus modality ranging from pictures to (nonsense) words and auditory stimuli (e.g. Lane et al. 2007; Speelman et al. 2013; Spruyt et al. 2011; Vande Kamp 2002). In the RRT, by

contrast, both inducer stimuli and statements are presented in a written form (De Houwer et al. 2015). This means that if one wants to use auditory samples of a language (variety) or linguistic variable, an alternative solution has to be found to include them in the experimental design outside of the stimuli, if the RRT is the preferred method (see for instance Chapter 4). As for the length of the stimuli, there are strict limitations in all three methods. Given that speed is an important component of automaticity, using short stimuli is critical to enhance the implicit character of the methods: the longer the stimuli, the more time participants have to process them, the less automatic their reactions. Hence, social psychologists tend to use pictures and short written words in AP or the IAT. If one wants to use auditory stimuli, they should be short as well. Speelman et al. (2013), for instance, use short words of approximately 600 ms in their AP experiment. In the RRT, statements are used as stimuli, which makes for longer stimuli. However, De Houwer and colleagues (2015) argue that this does not mean the RRT cannot measure processes that operate under conditions of automaticity. In the RRT, participants are also instructed to react to the stimuli as quickly as possible and in that sense still have a lot less time to reflect than in for instance a traditional questionnaire. For a further discussion of the RRT and implicitness, we refer to De Houwer et al. (2015).

1.C USE OF LINGUISTIC STIMULI | A third component of flexibility relates to the two previous ones (stimulus modality and stimulus length) and is of particular importance for linguists: can linguistic stimuli be used in AP, the IAT and the RRT? As demonstrated by Speelman et al. (2013) and for instance Redinger (2010), Pantos & Perkins (2012) and Campbell-Kibler (2012), AP and the traditional IAT seem to work well with linguistic stimuli (auditory or written), be it with the limitation of very short stimuli[15]. Note that this length restriction in AP and the IAT has implications for the type of linguistic phenomena that can be straightforwardly implemented in these measures. For phenomena requiring longer stimuli (like certain intonation contours or

[15] To make this more concrete: as mentioned above, Speelman et al. (2013) use auditory stimuli of maximum two syllables and approximately 600 ms. The stimuli used in the studies from Chapters 2 and 3 are similar in length. Pantos & Perkins (2012) report using auditory stimuli of up to 12 words (ca. 24 syllables).

syntactic patterns) other measures may be a more suitable choice. Additionally, if one is interested in measuring attitudes towards a variety or language as a whole, the variety will have to be represented by a small and carefully selected set of stimuli (see also criterion 2.B). No other variants of the IAT other than the traditional one have been used with linguistic stimuli and neither has the RRT. Chapters 2 and 3 will explore the use of linguistic stimuli in the P-IAT and chapter 4 reports a linguistic RRT study.

1.D STRUCTURE OF THE ATTITUDE OBJECT | In 2.2.3, it was explained that the standard IAT requires a binary target and attribute concept. In some cases this binary structure may be a restriction, for instance when the aim of the study is to measure associations either with a single attitude object or with more than two attitude objects. In the latter case, multiple IATs will have to be included in the experimental design (see criterion 5 for the consequences regarding within vs. between participant designs). However, as will also be explained in Chapter 2, a binary structure should not necessarily be regarded as a restriction. We believe that evaluation is inherently comparative, and in that sense, specifying the second component of the comparison gives a more controlled experimental design and enhances the experiment's ecological validity.

Variants of the IAT have been developed that allow researchers to measure attitudes towards a single attitude object (Single Category IAT, SC-IAT, Karpinski & Steinman 2006; or Single Target IAT ST-IAT, Wigboldus et al. 2004) or based on a single rather than binary attribute category (Single Attribute IAT, SA-IAT, Penke et al. 2006). A problem arises, though, when one wants to use these IAT variants with auditory linguistic stimuli. In these versions of the IAT, it is advisable to use target and attribute stimuli of the same modality in order to avoid recoding (Gawronski et al. 2011). What could happen is that participants use the modality of the stimulus as the criterion to categorise it rather that the intended target and attribute concepts. In that case the IAT is not measuring what it intended to measure anymore. Take the example of a ST-IAT used to measure positive/negative attitudes towards a certain accent represented by auditory stimuli. To avoid recoding in this example, the attribute stimuli would also have to be auditory. The question then is which accent to use for the attribute stimuli. If the same accent is used for

the target and attribute stimuli, the distinction between the two is not clear enough anymore for the IAT to function properly. Yet, if another accent is used for the attribute stimuli, a confound would be introduced into the design. One could argue that a neutral accent can be used for the attribute stimuli, but it may obviously be questioned whether there is such a thing as a neutral accent.

AP, by contrast, is somewhat more flexible, allowing more than two language varieties within a single experiment (e.g. Speelman et al. 2013). However, as explained in Section 2.2.2, it is advisable to include a neutral baseline category in AP experiments (Wentura & Degner 2010). For certain studies, the same issue will arise as the one raised for the single concept variants of the IAT above: it may not always be straightforward to identify a neutral category (Lane et al. 2007; Wentura & Denger 2010). As for measuring attitudes towards a single concept with AP: the minimum number of categories required for AP remains two (one of which could be a baseline category), given that the technique relies on the comparison of reaction times. The RRT is the method out of the three that will most straightforwardly allow single attitude objects (as well as binary categories): the statements can be formulated in an absolute or comparative manner. For instance, the RRT can include statements of the type 'variety A sounds intelligent', or of the type 'variety A sounds more intelligent than variety B'. This makes it possible to include both single and binary attitude objects in the RRT.

1.E TYPE OF EVALUATION | A way in which all three methods prove highly flexible is the type of evaluation they measure. The techniques are not restricted to measuring attitudes in terms of positivity and negativity, but can include any other type of qualification of the attitude object (e.g. Spruyt et al. 2011; Gawronski & De Houwer 2014; De Houwer et al. 2015; Koning et al. 2016; Dewitte et al. 2017). This makes the three measures particularly interesting for linguistic attitude research which usually focuses on dimensions like prestige or social attractiveness rather than just valence (see Section 1). In this context, we would like to point out that AP, the IAT and the RRT are methods suitable to investigate a specific hypothesis, but less appropriate for explorative research. The reason is that the type of evaluation (e.g. prestige or social attractiveness) included in the experiment needs to be selected by the

researcher beforehand (cf. the MGT in Section 1). This selection can be made based on theory or findings reported in the literature, but also by conducting preliminary studies using some of the more traditional attitude measures, like the 'draw a map' task, which are better suited to *explore* how attitudes are structured in a speech community. Note also that the number of attitudinal dimensions that can be included in a single AP, IAT or RRT experiment is restricted to one. This is an important difference with traditional methods like the MGT, where the researcher has the opportunity to include a wide variety of rating scales that may represent multiple attitudinal dimensions.

1.F POSSIBILITY TO IMPLEMENT CONTEXT | As will be discussed at length in Chapter 3, AP, the IAT and the RRT decontextualise attitude objects to a large extent. These attitude objects are represented by labels and/or a small set of short stimuli without any further contextual information. To better understand the social meaning of a language variety or variant in a specific context, it is necessary to take that context into account in the experimental design (Soukup 2015). Of course, a controlled experimental approach has practical limitations regarding the inclusion of contextual information, but we think the researcher should try to include as many context cues as the experimental design will allow in order to increase the method's ecological validity. Both AP and the IAT allow certain types of contextualisation. For the IAT, we refer to the discussion in Chapter 3. For AP, we will briefly illustrate one approach of including context cues in the experiment. Maddux and colleagues (2005) successfully manipulated attitudes towards Black vs. White people by using background pictures during the trials that evoked situations typically associated with prejudice towards one race or the other. The RRT has only been introduced very recently and, to the best of our knowledge, no studies have been published so far that explore the context sensitivity of the measure. It may be worth experimenting with the inclusion of context cues in the statements of the RRT, or explore the use of background pictures as context cues in the measure like in Maddux et al. (2005).

1.G SPECIFICATION OF THE ASSOCIATION BETWEEN THE ATTITUDE OBJECT AND ITS QUALIFICATION | The penultimate aspect of the flexibility criterion relates to whether the measures specify the relation between the attitude object and its

qualification. As was explained in Section 2.2.4, the RRT uses propositions that specify that relation. AP and the IAT do not, although the personalized variant of the IAT (the P-IAT) does specify the relation to a certain extent by using attribute labels like 'I like/don't like' or 'I want/ don't want'.

1.H LAB VS. ONLINE TESTING | A final aspect of flexibility concerns the physical location where the experiments can be conducted: lab conditions vs. online studies. AP, IAT and RRT experiments tend to be administered in lab conditions where the researcher can make sure respondents understand the instructions correctly and where sources of distraction are removed (which is crucial when measuring reaction times). Despite improved control over factors like these, conducting experiments in a lab context makes it more difficult to recruit participants as they have to make the effort to come to the lab. Online versions of experiments, on the contrary, allow an experimenter to reach a larger and perhaps more diverse pool of respondents in a shorter period of time (Friese et al. 2007). However, when participants take part in an experiment online, the researcher loses all control over the circumstances of the experiment (Glashouwer et al. 2013). Sometimes control questions like 'were you disturbed at any point?' are included in online studies to filter out respondents who were distracted during the experiment (Friese et al. 2007). In the end, there will always be a trade-off between collecting large amounts of data cost-effectively in a short time span, and conducting experiments in a highly controlled setting. The IAT is regularly used in an online version (e.g. Greenwald et al. 2003; Friese et al. 2007; Xu et al. 2014). To the best of our knowledge, online use of AP is not as common as it is for the IAT. The RRT research published so far has been lab-based, yet our RRT study in Chapter 4 was internet-based.

2.A NECESSITY OF LABELS | The second criterion in this review concentrates on the use of **labels** and their interaction with the **stimuli** in the experiment. Both the IAT and the RRT require the attitude object to be labelled. The former uses labels as reminders of the mapping of the target/attribute categories onto the response keys in the top corners of the screen. The latter relies on statements that contain propositions about the attitude object as stimuli. In order to be able to say something about the attitude object in these statements, some label will have to be chosen to refer to it. AP, by contrast, does

not require the attitude object to be labelled. The disadvantage of having to label the attitude object is that the choice of a label may prove to be difficult. For example, one may be interested in studying the social meaning of a linguistic variable for which laymen have no terminology. Or it may be unclear how non-linguists refer to a certain language variety, that is if they have a term to label the variety under study at all. And even if a label is available, a further problem may be that it is uncertain whether all respondents understand the label in the same way. Furthermore, linguistic attitude research has shown that the choice of a label for a variety may impact the attitudes measured (Coupland & Bishop 2007, see also Chapter 4). In that respect, it may be considered an advantage of AP that no labels have to be selected. However, without a label (in case an unambiguous one can be found) and thus solely relying on (audio) samples of a language variety, one cannot be certain that participants identify the samples as the language variety the researcher intended them to represent. This issue has also been raised as a criticism toward the MGT (Garrett 2010).

2.B IMPORTANCE OF CONCEPTS (LABELS) VS. EXEMPLARS (STIMULI) | A question closely related to the above: what is the influence of concepts (labels) and exemplars (stimuli) on the attitudes measured in AP, IAT and RRT experiments? Given that the attitude object is represented exclusively by exemplars in AP and by concepts in the RRT, this question is mainly relevant for the IAT which includes both a conceptual representation of the attitude object and one through exemplars. Previous studies have found that both labels and stimuli matter: it is the way in which the labels and stimuli jointly construe the attitude object that determines the attitudes measured (Lane et al. 2007). For example, Mitchell and colleagues (2003), conducted an IAT experiment in which they manipulated the category labels. They used the same exemplars of popular Black and unpopular White people (athletes and politicians respectively) and found that respondents had more positive attitudes towards the Black people when these people had to be categorised based on professional occupation than when they were categorised based on race. In the same study, Mitchell et al. (2003) report that the choice of stimuli can alter results of an IAT as well. When well-liked Black people represented the category 'Black', participants' preference for White people diminished compared to an IAT

where the category 'Black' was represented by disliked Black people. This means both labels and stimuli have to be selected carefully when designing an IAT experiment. Additionally, stimuli should be chosen so they are easy to categorise for respondents (Lane et al. 2007). If this is not the case, it may lead to extra unwanted error variance in the collected data. In addition to choosing stimuli that are highly representative of the attitude object, categorisation can be made easier for participants if additional cues are included in the experimental design that help distinguish between target and attribute stimuli. For example, one can use a different modality for target and attribute stimuli (e.g. words vs. pictures) or a different colour for words that constitute either type of stimuli (Lane et al. 2007). Although influence of both category labels and stimuli have been reported for the IAT, some studies suggest the impact of the category level is stronger (Lane et al. 2007: 84).

3. PSYCHOMETRIC QUALITIES | Our third review criterion is **psychometric quality**. This concerns a method's validity and reliability. A valid method is one that measures what it set out to measure. So for an implicit attitude measure to be valid, the effects it captures should reflect variations in the targeted attitudes. Reliability refers to a method's consistency: a reliable method should produce the same results under the same circumstances. The validity and reliability of a method depend on the specifics of a study (which attitudes are being measured and what procedural choices have been made). Given that AP, the IAT and the RRT are frameworks used to assess a wide variety of attitudes and beliefs and given that there is a lot of variation in their procedural implementation, one has to be careful with generalizations about their psychometric qualities (Lane et al. 2007). However, considering evidence from a large range of studies, the IAT is deemed to have good psychometric qualities (Lane et al. 2007; Gawronski et al. 2011). Generally, psychologists assume that the IAT effects are caused by the attitudes the measure tries to capture (De Houwer et al. 2009). Hence, it is seen as a valid measure. Reliability for the IAT in its traditional form is also satisfactory (internal consistency is usually situated between .70 and .90, Gawronski & Hahn in press). The P-IAT, the IAT variant used in Chapters 2 and 3, approaches the traditional IAT in its reliability (Gawronski et al. 2011).

Given that the RRT has only been developed very recently, the amount of published research based on the measure is still limited. As a result, it is even more difficult to make generalizations about this method's psychometric qualities. However, the studies published so far indicate acceptable reliability estimates (De Houwer et al. 2015; Koning et al. 2016). De Houwer and colleagues (2015) also started investigating the measure's validity by studying its correlation with other attitude measures. Although further research is needed, their results seem promising.

The psychometric qualities for AP are less positive: often scores of Cronbach's alpha below .50 are reported (Gawronski & De Houwer 2014). This is problematic, because internal consistency scores this low make it difficult to measure the effect of an experimental manipulation, but also to capture correlations with other measures (Gawronski & Hahn in press). In addition to low internal consistency, test-retest reliability is problematic for AP: often little to no correlation is found between repeated AP experiments (De Houwer et al. 2009). Despite unsatisfactory reliability, AP is assumed to be a valid attitude measure, although more research would be desirable to further substantiate this assumption (De Houwer et al. 2009). For a detailed discussion of the available evidence for reliability and validity of the IAT and AP, we refer to De Houwer et al. (2009). For the RRT, we refer to De Houwer et al. (2015).

4.A TASK LENGTH | In the fourth criterion of our review, we focus on potential **participants** and how taking part in studies involving one of our three implicit measures may affect them. Firstly, an important practical aspect of any method is how long participants need to complete the task. This is, of course, highly dependent on the specifics of the experimental design and on the individual participant (e.g. how long they need to read instructions or how much additional information they ask for during the debriefing stage at the end of the experiment). Based on our own experience from conducting the studies reported in the following chapters, we estimate that the completion of one IAT or RRT easily takes up 10 to 20 minutes. AP experiments have been reported to take about the same amount of time (e.g. Speelman et al. 2013 report a full AP experiment to take around 15 minutes). If the implicit measurement is combined with additional data collection (e.g. demographic information about

respondents or an explicit attitude measure) or with a second IAT/RRT/AP experiment[16], the time required for the study quickly increases. This makes it more demanding for participants and harder for the researcher to find willing respondents or funds to recompense them for their efforts. Additionally, if the study is conducted in a lab context rather than online (cf. 1.H LAB VS. ONLINE TESTING in Table 1 and above), the study can become highly time-consuming.

4.B COGNITIVE LOAD | Another aspect to consider is how cognitively demanding AP, the IAT and the RRT are for participants. All three measures require participants' utmost concentration and fast reactions during ca. 15 minutes. The measures are also computer based, so some familiarity with computers will certainly facilitate the correct completion of the experiment. As became evident from the procedural details in Sections 2.2.2 - 2.2.4, the IAT and RRT are perhaps even more complicated for participants than AP in certain respects. In the latter, there is only one task participants have to become familiar with and one type of stimuli they have to categorise. All these elements combined may make an AP/IAT/RRT experiment rather demanding for certain groups of participants, despite the fact that at least the IAT has been successfully used with less obvious audiences such as pre-school children (given the necessary procedural adaptations, Cvencek et al. 2011). It should therefore be carefully considered whether the method is suitable for the target audience of the study.

5. REPEATED MEASUREMENTS | A fifth criterion which may be of use in determining whether the measures under review are suitable for one's research is the potential for **repeated measurements**. As explained above, there are restrictions on the number of attitude objects or types of evaluation that can be included in a single AP, IAT or RRT experiment. If one's research question pertains to multiple attitude objects (e.g. language varieties) or social meanings (e.g. prestige vs. social attractiveness), multiple versions of the experiment will have to be constructed. In that case, the question arises whether it is possible to work with within subject designs where one participants completes two or

[16] Given that the number of attitude objects are restricted in these methods (cf. supra), it is often necessary to include more than one AP, IAT or RRT experiment in a study. But see also criterion 5 regarding repeated measures.

more IATs/RRTs/AP experiments. If not, a between subject design has to be used, which means the number of participants required for the study will increase quickly. It is not uncommon to have participants complete multiple successive AP experiments, IATs or RRTs (e.g. Scherer & Lambert 2012; Friese et al. 2007). However, there is always a risk of fatigue effects: as explained above, these implicit measures require a lot of concentration from participants (e.g. Spruyt et al. 2007; Bar-Anan & Nosek 2014). Additionally, when respondents perform multiple tasks, the possibility of practice effects should be taken into account (Fiedler & Bluemke 2005; Gawronski et al. 2011; Bar-Anan & Nosek 2014).

6. TYPE OF MEASUREMENT: ABSOLUTE VS. RELATIVE | Our penultimate criterion in this review concerns the **type of measurement**: does the method allow the measurement of absolute attitudes towards an attitude object, relative attitudes comparing two (or more) attitude objects, or both? As a result of its binary structure, the IAT is a relative attitude measure: it aims to capture attitudes towards two attitude objects relative to one another without reference to a neutral benchmark (Lane et al. 2007)[17]. AP is also inherently relative: in order to measure whether participants react faster of slower to a target stimulus after a certain category of primes, there needs to be a second category or a baseline to allow comparison of reaction times (Wentura & Degner 2010). Yet, when using AP, it is possible to calculate a separate priming score for the positive and negative associations with each of the attitude objects included in the experiment (compared to the neutral baseline). This is impossible for the IAT where both associations are conflated into one score (Gawronski & De Houwer 2014). The RRT, by contrast, offers the option of both absolute and relative measurement. It is possible to formulate the statements and rules used in this measure in an absolute or relative way (e.g. 'variety A sounds prestigious' or 'variety A sounds more prestigious than variety B'). The fact that the RRT can function in both ways may make it the most attractive option in cases where a single attitude object is studied, but we would like to stress that the relative

[17] There are variants of the IAT which allow for absolute attitude measurement, such as the ST-IAT. Yet, as explained in (1.D), these are not easily implemented if one wants to use auditory linguistic stimuli.

nature of the IAT and AP should not necessarily be viewed as a limitation. As will be discussed in Chapter 2, we believe that evaluation of language variation is an inherently comparative process: a language (variety/ feature) is always judged relative to another language (variety/feature) that was not used in a certain context. In that sense, making the comparison explicit in the attitude measure leads to a better controlled experimental design and increased ecological validity.

7. WHAT RESPONSES ARE RECORDED? | The final characteristic of AP, the IAT and the RRT we will discuss here pertains to the **type of responses** that are measured. For all three techniques, both reaction times and the correctness of the response are recorded. So regarding this criterion, the measures are not distinguishable. What we would like to highlight here are the consequences of this characteristic, especially in comparison to most of the traditional language attitude measures that do not rely on response latencies (e.g. the MGT or 'draw a map' task). Accurate reaction time measurement requires good computers and reliable software[18], and is more prone to technical failure than techniques relying on pen and paper. In addition, measures based on response latencies require circumstances that cancel out any potential source of distraction. As a result, it is desirable that participants complete the experiment individually. This means that conducting studies based on reaction times may be less time-efficient than for instance a MGT experiment or 'draw a map' task which can easily be administered in group.

[18] A wide variety of software is available to programme AP, IAT or RRT experiments. Both commercial (e.g. E-Prime) and open source (e.g. Affect 4.0, Spruyt et al. 2010) software are commonly used.

CRITERIA			AP	IAT	RRT
1. FLEXIBILITY	*A*	*stimulus modality*	flexible (words, pictures, sound,…)	flexible (words, pictures, sound,…)	written propositions
	B	*stimulus length*	short	short	somewhat longer
	C	*use of linguistic stimuli?*	yes	yes	? to be tested (cf. Chapter 4)
	D	*structure of the attitude object*	2(+) (incl. neutral baseline)	binary	1(+)
	E	*type of evaluation*	positive-negative + other (semantic priming)	positive-negative + other	positive-negative + other
	F	*possibility to implement context*	yes	yes	? to be tested
	G	*specification of the association between the attitude object & qualification*	no	no (except P-IAT)	yes
	H	*lab vs. online testing*	less frequently used online	frequently used online	possible online (cf. Chapter 4)

2. LABELLING & STIMULI	*A*	*labels necessary?*	no	yes	yes
	B	*importance of concepts (labels) vs. exemplars (stimuli)*	exemplar based	concept & exemplar based	concept based
3. PSYCHOMETRIC QUALITIES			reliability unsatisfactory	good reliability	good reliability
4. STRAIN ON PARTICIPANTS	*A*	*task length*	ca. 15 mins (depending on study and individual participants)		
	B	*cognitive load*	all require considerable concentration; IAT perhaps most complex procedure, AP least complex instructions		
5. REPEATED MEASUREMENTS			yes, but possibility of fatigue/practice effects		
6. TYPE OF MEASUREMENT: ABSOLUTE VS. RELATIVE			relative	relative	relative or absolute
7. WHAT RESPONSES ARE RECORDED?			reaction times + correctness response	reaction times + correctness response	reaction times + correctness response

Table 1 - Summary of the comparative review of AP, the IAT and the RRT

Based on the comparative review above, can we come to a preliminary evaluation about the potential of AP, the IAT and the RRT for linguistic research on the social meaning of language variation? And if that evaluation is positive, can we select one of the three methods as most suitable or most promising for linguistic attitude research? Let us start with the first question. After successful implementations of AP, the IAT and the RRT in a diverse range of studies in their field of origin, the first linguistic applications of the IAT (e.g. Redinger 2010; Pantos & Perkins 2012; Campbell-Kibler 2012) and AP (Speelman et al. 2013) seem encouraging. Yet, we think some caution is called for before sociolinguists start using these measures on a large scale in their research. Firstly, it is important that researchers are well aware of the qualities – the potential, but also the limitations – of the measures. There are considerable technical aspects to get right and procedural choices to be made which we deem should be considered carefully in order to produce high-quality experiments. Given a good command of the practical and technical aspects of these new attitude measures, we think they have potential to become a useful and complementary addition to the existing range of methods.

Yet, even if one is well-informed about the procedural and technical side of these implicit measures, we believe caution is still required: there are aspects of AP, the IAT and the RRT that are not yet fully understood. Psychologists do not fully agree or have multiple alternative theories about which psychological processes underlie measures like AP or the IAT and to which extent the effects captured by the measures are caused by the attitudes under study or by other task related aspects[19] (Lane et al. 2007; Gawronski et al. 2014). Partially because those underlying processes are not fully understood, the impact of certain procedural variations has no univocal explanation yet. Also, despite elaborate theorising about the concept of implicitness, not all measures deemed implicit have been submitted to extensive empirical investigation regarding some aspects of their implicitness (De Houwer et al. 2009).

---

[19] Psychologists are using mathematical modelling to try and tease apart the contribution of different processes to the effects measured by techniques like the IAT (Gawronski & De Houwer 2014). An overview of this line of research can be found in Sherman et al. (2010).

Before we can utilize these methods to their full potential, we need more knowledge about the processes underlying them and about their implicitness. While some may consider this work to fall within the scope of social psychology, rather than sociolinguistics, linguists should also further develop their theorizing regarding language attitudes and be more explicit about their interpretation of certain concepts relating to language attitudes. For instance, how do we define implicit (or 'covert') language attitudes and how does this relate to the way psychologists interpret implicitness in terms of automaticity? What is the role of implicit language attitudes in processes of language variation and change? And if these attitudes do play a role, what aspects of implicitness are important (simply awareness, or also other automaticity components)? How do linguistic theories about the link between language and social meaning (e.g. exemplar theory, Foulkes & Docherty 2006, see Chapter 3) relate to the ideas behind these social psychological implicit attitude measures? Engaging with questions like these will lead to a better insight into the potential of the social psychological attitude measures discussed in this chapter, and will allow sociolinguists to make more confident decisions about the use of these methods in their research. Taking the inverse perspective, we also believe that interacting with the measures can provide an impetus for further developing sociolinguistic theory, because it may help to identify which aspects of sociolinguistic theory remain vague or which concepts need clarification. Examples of such aspects or concepts are – as mentioned above –implicitness and the relationship between the social meaning of language and language change, but also questions concerning categorisation and the cognitive representation of language features and varieties (as will be discussed in in the context of the study presented in Chapter 4). Formulating answers to these questions will again allow us to make more confident decisions about which methods are best suited for our research questions, to better understand how results from different methods relate to each other, but also to optimize the procedural details of the methods we use (Lane et al. 2007) In this way, we get a fruitful reciprocity between dealing with theoretical and conceptual issues on the one hand, and methodological exploration on the other. We will come back

to some of the questions raised here in Chapter 5 where they will be discussed in the light of the studies reported in Chapters 2, 3 and 4.

As for the second question raised in the context of the comparative review between AP, the IAT and the RRT – whether we can recommend one of the three measures as the most suitable or promising for linguistic attitude research, the answer is decidedly 'no'. The comparative review has made clear that all three methods have both benefits and limitations. For instance, AP offers the advantage of not having to select labels for the linguistic phenomena under study, yet the method has been shown to be less reliable than the IAT or RRT. Moreover, what is considered a benefit or limitation strongly depends on the specific goals of a study: the binary structure of an IAT, for example, may be ideal for a study on the comparison of attitudes towards two language varieties, but perhaps not for a research question concerning one variety only. Choosing the best methodology for a study thus depends on the specific research questions, the (stimulus) materials that are on hand, the characteristics of the linguistic phenomenon under study, the time and funds available to prepare the experiment, characteristics of the targeted respondents, the time within which the study needs to be conducted etc. These considerations are of course not specific to the social psychological implicit attitude measures considered in this chapter, but pertain to all carefully planned and designed experimental studies.

Chapter 2

Exploring the Personalized Implicit Association Test as a measure of language attitudes

1 Introduction

This chapter reports on a study that uses the Personalized Implicit Association Test (P-IAT, Olson & Fazio 2004) to measure attitudes towards regional language variation in Belgian Dutch. The study aligns with a recent wave of methodological innovation in the field of language attitude research, as well as with a renewed interest in attitudes towards (regional) language variation in the Dutch language area, i.e. The Netherlands and Flanders, the Dutch speaking part of Belgium (e.g. Grondelaers et al. 2011; Speelman et al. 2013; Preston 2016; see also Chapter 1). In that respect, the objectives of the study are twofold: both methodological and descriptive. From a methodological point of view, the study introduces the Personalized Implicit Association Test, an existing social psychological attitude measure, as a new method to measure language attitudes. From a descriptive perspective, it aims to contribute to the study of attitudes towards regional variation in Belgian Dutch, which has received very little attention compared to the work that is being carried out on

variation in Netherlandic Dutch. In what follows, we will introduce the study from both perspectives.

1.1 Methodological perspective

As explained in Chapter 1, quantitative linguistic attitude research has known little methodological innovation in the last few decades compared to social psychology, which has witnessed an explosion of new attitude measures in recent years (Grondelaers 2013; Speelman et al. 2013; Gawronski & De Houwer 2014). After a period of limited innovation, linguistic attitude research now seems to be catching up (Preston 2016). While some of this recent innovation in linguistics relates to the traditional methods used in language attitude research (see Section 2.1 in Chapter 1), some linguists have started to explore attitude measures developed in social psychology. In Chapter 1, we mentioned that these explorations predominantly focus on the IAT (Babel 2010; Redinger 2010; Pantos & Perkins 2012; Lee 2015; Watt & Llamas 2015; Loudermilk 2015; Leinonen 2016; McKenzie 2017), with the exception of Speelman et al. (2013) who implemented affective priming in their language attitude research. The dominance of the IAT in these linguistic explorations of new attitude measures, reflects the method's popularity in social psychological research, which is due to a number of advantages the IAT entails. For a detailed discussion of some of those, we refer to the preceding chapter.

As became clear in that discussion in Chapter 1, the IAT also comes with some downsides, one of which is the measurement of extra-personal associations instead of personal associations (although this claim is not uncontroversial, Gawronski et al. 2008). Personal associations are those endorsed by an individual. Extra-personal associations, by contrast, are present in an individual's memory, because he or she has come across these associations, but does not necessarily endorse them. If we take the example about vegetables from Chapter 1: even when an individual dislikes vegetables, a traditional IAT could still return favourable attitudes towards these foods, because that person has undoubtedly been in contact with campaigns in the public domain teaching people that vegetables are good for them. The Personalized IAT (P-IAT) (Olson & Fazio, 2004) is an adaptation of the

traditional IAT that aims to avoid measuring extra-personal associations. The defining features that personalize an IAT are the use of attribute labels that refer directly to the participant's opinions, such as 'I like' and 'I don't like' instead of 'good' and 'bad', as well as leaving out corrective error feedback. In this study, the personalized variant of the IAT will be applied for the first time – to the best of our knowledge – in the context of linguistic attitude research. We will show that the P-IAT is a promising addition to the range of methods used in linguistics to study the social meaning of language variation. We will do so by comparing the P-IAT to two other measures: one targeting implicit attitudes (AP) and one explicit attitudes (a rating task using semantic differential scales).

1.2 Descriptive perspective

Since the 1980s, attitude research towards (regional) varieties of Dutch has mainly focused on Netherlandic Dutch (Grondelaers 2013). Even after 2000, when language attitude research in the Low Countries slowly started to increase again after diminished interest in the 1990s and focused its attention on variation in the standard language (Grondelaers 2013), most studies tend to concentrate on Dutch variation in the Netherlands (e.g. Van Bezooijen 2001, Grondelaers & Van Hout 2010). Recent attitude research towards variation in Belgian Dutch is still scarce and mainly focuses on colloquial Belgian Dutch (CBD), an informal variety of the standard language with a regional flavour, and its relation to the standard variety and local dialects (Grondelaers 2013). Overall, this research is rather fragmented, focussing on various regional varieties and participant groups. A large-scale survey of the attitudinal landscape in Dutch speaking Belgium is lacking as yet.

As mentioned, most research on attitudes towards variation in Belgian Dutch focuses on CBD or *tussentaal*. This highly variable colloquial variety, which has proven hard to define and delineate, is situated somewhere between the local dialects and Standard Belgian Dutch (SBD). This explains why the variety is also nicknamed *tussentaal,* which literally translates as 'in-between language' (Grondelaers et al. 2011). CBD can be described as 'a collection of linguistic variables that have a supra-regional distribution on the geographic

dimension' (Geeraerts & Van de Velde 2013: 532). These variables are phonological, morphosyntactic and lexical in nature. Many of them find their origin in the central Brabant area of Belgium, which is perhaps not surprising given the dominant role of this region in the linguistic history of Belgian Dutch (Goossens 1970; Geeraerts & Van de Velde 2013). However, CBD often also includes regional elements, such as lexical items and a regional accent (Geeraerts & Van de Velde 2013).

The few recent attitudinal studies carried out in Dutch speaking Belgium, all dealing with CBD compared to SBD[20], mostly come to the conclusion that SBD is viewed more positively than CBD, specifically regarding perceptions of power, competence and status (Cuvelier 2007; Impe 2006; Impe & Speelman 2007; Speelman et al. 2013). The amount of CBD features included plays a mediating role in this trend: the more features, the less status (Impe & Speelman 2007). CBD, by contrast, is perceived more positively on the social attractiveness dimension (Cuvelier 2007; Impe & Speelman 2007). Yet, some studies report findings that nuance this picture and present some neutral (Lybaert 2014), less negative (Cuvelier 2007; Grondelaers & Speelman 2013) or inconsistent (albeit rather low, Vandekerckhove & Cuvelier 2007) perceptions of CBD on the level of competence and status. Grondelaers et al. (2011), who did not include a comparison with SBD in their study, even report a certain level of speaker prestige and accent status for CBD, especially for the central Brabant variety in comparison to more peripheral varieties. The influence of sociodemographic variables on these attitudes towards SBD and CBD is still unclear. For instance, certain studies found (some) influence of listeners' age (Vandekerckhove & Cuvelier 2007; Ghyselen 2009), while others do not (Impe & Speelman 2007, who did not find any gender differences either). Impe & Speelman (2007) and Grondelaers et al. (2011) also report no influence of listeners' regional background, while this does seem to be the case

[20] Two exceptions need to be mentioned. One is Grondelaers et al. (2011) who did not include SBD in their study and focused on a variety in between fully-fletched CBD and SBD. They measured perceptions towards Dutch spoken by teachers which is regionally accented and contains some CBD features, but is also fairly close to SBD. A second exception is Marzo (2016) who studies attitudes towards *Citétaal*, an urban vernacular mainly associated with the province of Limburg in the east of Flanders.

in Speelman et al. (2013). It has also been shown that the regional origin of the CBD variety influences perceptions: in Impe & Speelman (2007), the Brabant variety of CBD receives the most positive evaluations on the social attractiveness dimension. It is important to keep in mind though, that the methods, designs, and the investigated varieties of CBD differ between these studies, so a direct comparison of results is difficult.

A complementary perspective with regard to these findings is provided by Van Gijsel and colleagues (2008). In their study, the use of CBD and SBD in Belgian radio and television commercials is investigated. Perceptions towards both varieties turn out to be deliberately exploited in advertisements: not only are commercials containing CBD usually directed towards a younger audience, there is also a division of labour between both varieties. CBD tends to be used for staging informal everyday conversations, while serious and factual information is delivered in SBD. These findings from production research seem to correspond with the ones obtained in the perception studies on CBD and SBD.

All perception studies mentioned above take a more holistic perspective and study attitudes towards CBD without distinguishing between different types of CBD features, except for Grondelaers & Speelman (2013) and Speelman et al. (2013). The former takes into account phonological, morphological and lexical features, while the latter focuses on regional pronunciation. Grondelaers & Speelman (2013) found that evaluations of CBD differ depending on the nature of features presented to the listener-judges. CBD lexis and morphology are both downgraded on the prestige dimension, and so are morphological features for dynamism (a dimension not taken into account by other studies). Yet, CBD phonology is not downgraded on either prestige or dynamism and CBD lexis is even upgraded on the latter dimension.

Although the study reported in this chapter does not allow us to distinguish between different dimensions of language attitudes (see Chapter 4 for a study that does), we will confirm the positive evaluation of SBD compared to CBD reported by previous work. We will also be able to demonstrate that the interaction between regional origin of the participants and the variety of CBD is of relevance to the language attitudes under study.

1.3 Research questions and hypotheses

Against the methodological and descriptive background sketched in Sections 1.1 and 1.2, we can now specify the aims of the present study with regard to both perspectives. From a methodological point of view, we explore the P-IAT as a measure of language attitudes and show how the measure can be a useful tool for linguists. We opted for the personalized variant of the IAT, because it has been demonstrated to reduce the risk of measuring extra-personal associations while still sporting good psychometric qualities comparable to those of the traditional IAT (Gawronski et al. 2011). Additionally, we aim to compare the performance of the P-IAT as a measure of language attitudes to affective priming (AP). The latter method has been successfully applied to measure language attitudes by Speelman and colleagues (2013). Yet, social psychologists have shown that AP does not do so well psychometrically, specifically when it comes to reliability (Spruyt et al. 2011). Hence, in this study we set out to explore whether we can obtain similar results using the more reliable P-IAT. In order to do so, we applied the P-IAT to study the same regional varieties of Belgian Dutch that were investigated in Speelman et al. (2013) using identical stimuli to guarantee maximal comparability between the two studies. Additionally, we collected explicit ratings about the language varieties under study, so the P-IAT results can be compared with these as well. As the results will show, the observed attitudinal patterns largely coincide between the three measures, but are not identical. In the discussion in Section 4, we will consider a number of potential explanations why there is no perfect overlap.

From a descriptive point of view, this study aims to contribute to mapping out the attitudinal landscape of Belgian Dutch, a task which is far from complete. We measured attitudes towards SBD and two regional varieties of Belgian Dutch, one central variety and one peripheral variety. The choice of specific regional varieties, Antwerp as the central variety and West-Flemish as the peripheral variety, was based on Speelman et al. (2013) in order to be able to compare results. Choosing a central and a peripheral variety is also interesting from a theoretical point of view: CBD features from the central area are claimed to spread to the peripheral areas, and, as indicated above, perceptual research found some evidence that central varieties may be more

positively evaluated than peripheral ones (Impe & Speelman 2007; Grondelaers et al. 2011).

Our study focuses on regional accent, which is an important feature of CBD varieties. The reason we decided to study regional accent in isolation is both theoretically, practically and descriptively motivated. Firstly, apart from Speelman et al. (2013) no recent work on the attitudinal landscape of Belgian Dutch has focused purely on the evaluation of accent. Yet, as Grondelaers & Speelman (2013) found, CBD features on different linguistic levels may carry different social meaning and for regional accent variation, this is virtually unexplored. Secondly, accent variation may be the most obvious type of variation to implement in the P-IAT as the measure requires its stimuli to be as short as possible. Hence, accent variation presents itself as a good starting point for exploring the P-IAT as a language attitude measure. Even so, we hope future research will experiment with the possibilities of including, for example, lexical and syntactic variation in the IAT paradigm (to the extent that the method allows this, see Section 4.3 below). Thirdly, the participants that took part in our experiments came from the central Antwerp area and peripheral West-Flanders. Choosing these two groups guarantees comparability with the Speelman et al. (2013) study, but it also allows us to further investigate whether language attitudes show regional stratification on the side of the listeners as reported in previous work (Impe & Speelman 2007; Grondelaers et al. 2011).

If the P-IAT measures attitudes in a similar way as AP, we expect to find the following pattern in the data, based on what was reported by Speelman and colleagues (2013): all participants prefer the standard variety and their own regional accent over the other group's regional variety. However, participants from the central Antwerp area are more positive about their own regional accent than about SBD, while the opposite is true for participants from the peripheral area. These hypotheses are summarised in Table 2.

Hypotheses based on Speelman et al. (2013)		
participant group	Antwerp participants (centre)	West-Flemish participants (periphery)
hypothesized attitudes	Antwerp variety (own) > SBD > West-Flemish variety (other)	SBD > West-Flemish variety (own) > Antwerp variety (other)

Table 2 - Hypotheses for the study in Chapter 2

2 Method

The study consisted of two parts: a P-IAT and a questionnaire. After giving informed consent, participants were asked to take part in a P-IAT experiment measuring implicit attitudes. This indirect attitude measurement was followed by a direct one using semantic differential scales and a short questionnaire collecting basic demographic information. Afterwards, participants were fully debriefed about the aims of the research project and the experiment they took part in.

2.1 Participants

In total 192 participants were recruited at a university campus in Kortrijk, West-Flanders, and at a university campus in the city of Antwerp. We decided to use university students as participants in an attempt to introduce a certain level of control over age and social background as these factors are known to have a potential influence language attitudes (e.g. Ghyselen 2009 for age; Impe & Speelman 2007 and Vandekerckhove & Cuvelier 2007 for gender). All participants originated from West-Flanders or Antwerp and were still living there. No psychology or linguistics students were allowed to take the experiment to avoid participants with previous experience with either the method or the topic of the study.

Data from 14 participants had to be discarded, because they came from regions other than West-Flanders or Antwerp. Additionally, 2 outliers were removed from the dataset. Despite the fact that these outliers satisfied the requirements to take part in the study, leaving their data in the analysis influenced results so that effects became significant which otherwise were not. Hence we deemed it justifiable to remove these participants from the dataset. Data obtained from these two students is excluded from all results reported below. Of the 176 participants included in the analyses, 102 were male and 74 were female with an average age of 20 (*SD* = 1.79, *MIN* = 18, *MAX* = 25).

2.2 P-IAT: Task, materials, procedure and design

The P-IAT measures the association between a binary target concept and a binary attribute concept by comparing reaction times in a series of categorisation tasks. The test used in this study was designed as described in Chapter 1, Section 2.2.3 with language variety (regiolect vs. SBD) as the target concept and valence (I like vs. I don't like) as the attribute concept. A schematic overview of the structure of the P-IAT as used in this experiment can be found in Table 3.

The reaction times recorded in a P-IAT are traditionally analysed using a scoring algorithm that produces so-called D scores, which are average difference scores between the reaction times in the experimental blocks of the P-IAT calibrated for variability in response speed between individuals (Greenwald et al. 2003). The variant of the algorithm used in this study is the D_6. We calculated D scores using the IAT package in R (Martin 2015) which is based on the algorithm described in Greenwald et al. (2003). The D scores were analysed using multiple linear regression. This method was deemed most robust given the slightly unbalanced design (see Table 4).

The stimuli and category labels used in the study are summarised in Table 5. All stimuli were selected from the stimulus set used in Speelman et al. (2013). The set of auditory target stimuli consisted of 6 neutral words, each produced in both regiolects and SBD by two professional male speakers matched for age and social background. Both speakers originated from and still lived in the region of which they produced the accent. They both have a clear

block	type of block	task	# trials	mapping left key	mapping right key	stimuli	stimulus mode
1	practice	target discrimination	20	Antwerp	SBD	6 SBD, 6 Antwerp accented words	
2	practice	attribute discrimination	20	I don't like	I like	5 positive, 5 negative pictures	
pre 3[21]	practice	target + attribute discrimination	4	Antwerp I don't like	SBD I like	6 SBD, 6 Antwerp accented words 5 positive, 5 negative pictures	
3	experimental	target + attribute discrimination	40	Antwerp I don't like	SBD I like	6 SBD, 6 Antwerp accented words 5 positive, 5 negative pictures	
4	experimental	target + attribute discrimination	40	Antwerp I don't like	SBD I like	6 SBD, 6 Antwerp accented words 5 positive, 5 negative pictures	
5	practice	target discrimination	40	SBD	Antwerp	6 SBD, 6 Antwerp accented words	
pre 6	practice	target + attribute discrimination	4	SBD I don't like	Antwerp I like	6 SBD, 6 Antwerp accented words 5 positive, 5 negative pictures	
6	experimental	target + attribute discrimination	40	SBD I don't like	Antwerp I like	6 SBD, 6 Antwerp accented words 5 positive, 5 negative pictures	
7	experimental	target + attribute discrimination	40	SBD I don't like	Antwerp I like	6 SBD, 6 Antwerp accented words 5 positive, 5 negative pictures	

Table 3 - Schematic overview of the structure of the P-IAT using experiment A from Chapter 2 (see Table 4) as an example

[21] These four trials, preceding blocks 3 and 6, contain a stimulus of each category to allow participants to get used to the double categorisation task. They are discarded from the analysis, so mistakes or shorter/longer RTs owing to unfamiliarity with the task will not unduly influence results.

and pleasant voice. The target stimuli were controlled for duration (*M* = 606.13 ms, *SD* = 29.58), length (two syllables), frequency (based on the *Corpus of Spoken Dutch*, Schuurman et al. 2003, and the *Football Corpus*), familiarity (based on ratings by 94 Belgian students), valence (based on ratings by 35 participants) and degree of colloquiality (measured through phonetic distance between the standard stimuli and the regional stimuli, see Impe (2010) for a detailed description). For the attribute stimuli, we used 5 positive and 5 negative real life colour pictures for which norm data have been collected (Spruyt et al.2002). All pictures were equal in size (410 x 308 pixels). The experiment was run on a laptop with a screen resolution of 1366 x 768 pixels using Affect 4.0 (Spruyt et al. 2010). For the auditory target stimuli, a Jabra UC VOICE 150 MS Duo headset was used.

The labels we selected for the attribute categories are *vind ik goed/ vind ik slecht* (literally 'I find it good'/ 'I find it bad'). This is the main feature that personalizes our IAT, in addition to leaving out corrective feedback for the attribute stimuli. For the target categories, the labels *Antwerps accent* ('Antwerp accent'), *West-Vlaams accent* ('West-Flemish accent') and *neutraal accent* ('neutral accent') were used. We chose not to label SBD as 'standard accent' to avoid normative associations as much as possible.

		A	B	C	D
	TARGET VARIETIES	*Antwerp SBD*	*Antwerp SBD*	*West-Flemish SBD*	*West-Flemish SBD*
	MAPPING RESPONSE KEYS	*order 1*	*order 2*	*order 1*	*order 2*
ORIGIN PARTICIPANTS	*Antwerp*	23	22	22	23
	West-Flanders	24	21	20	21

Table 4 - Between subject design of the implicit attitude measurement including participant numbers (Chapter 2)

The experiments were conducted individually in a quiet, dimly lit room. Participants were briefly informed about what was expected of them and signed a consent form if they agreed to participate. They were told that the experiment investigated how people process images and sound. After completing the P-IAT and the explicit attitude measurement, participant were fully debriefed.

The study comprised four P-IAT experiments which included pairings of SBD with each of the regiolects. The Antwerp variety was included in experiments A and B, while the West-Flemish regiolect featured in experiments C and D (see Table 4). The reason why two experiments were included for each pairing of SBD and one of the regiolects is that the IAT is known to suffer from block order effects: if the first set of experimental blocks are the congruent blocks then the IAT effect tends to be larger (Teige-Mocigemba et al. 2010; see also Chapter 1, Section 2.2.3). Because in this study we don't known in advance which is the congruent block for each participant and because it may not be same for all of them, we decided to counterbalance the order of the experimental blocks. In the analysis, results from experiments A and B will be pooled and treated as one experiment, and so will the results from experiments C and D.

	labels	Stimuli
target	*neutraal accent* ('neutral accent') *Antwerps accent* ('Antwerp accent') *West-Vlaams accent* ('West-Flemish accent')	*maken*[n,a,w] ('to make') *moment*[n,a,w] ('moment') *nummer*[n,a,w]('number') *pakken*[n,a,w] ('to get/grab') *snappen*[n,a,w] ('to grasp') *verstaan*[n,a,w] ('to understand')
attribute	*vind ik goed* ('I like') *vind ik slecht* ('I don't like')	5 positive real-life colour pictures 5 negative real-life colour pictures

[a] realised with an Antwerp accent; [n] realised with a SBD accent; [w] realised with a West-Flemish accent

Table 5 - Stimulus set for the P-IAT (Chapter 2)

In the between subject design, summarised in Table 4, each participant was randomly assigned to one of the four experiments. Because the IAT and its variants have been reported to suffer from practice effects (Gawronski et al. 2011), we decided to limit the number of P-IATs per participant to a single one.

2.3 Explicit attitude measurement and basic demographic information

After completing the P-IAT, participants were presented with two 10 point semantic differential scales. First they were asked to rate the regiolect that featured in their P-IAT. Next, they were presented with a scale to rate SBD. In order to ensure maximal comparability with the association measured in the P-IAT, the adjectives used on either side of the scale were Dutch equivalents of 'good' and 'bad' and the varieties were labelled in the same way they were in the P-IAT experiment. To mimic the personalized aspect of the P-IAT, the question was phrased as 'What do you think about an [Antwerp accent/West-Flemish accent/neutral (standard) accent]?'

The final element of the study before debriefing was a short questionnaire asking for basic demographic information (gender, age, region of origin, etc.). With the exception of participants' region of origin, this information was not collected to include in the analyses, but solely to be able to control the demographic background of the participants.

3 Results

3.1 Implicit attitude measurement

After computing D scores based on the reaction times measured in the experimental blocks of the P-IAT, we entered these difference scores into a linear regression analysis as the response variable with participants' region of origin and the language varieties included in the P-IAT as the predictor variables. The resulting model is summarised in Table 6. Note that sum coding was used, so the estimate for the intercept represents the grand mean.

The grand mean reported in Table 6 is significantly different from 0. Its positive value indicates that overall, there is a preference for SBD, given that positive D scores in this experiment represent a stronger association between SBD and liking. This trend is also clearly visible in Figure 4 which summarises the data per experiment. *T*-tests with Bonferroni correction confirm that both participant groups in all experiments significantly prefer SBD compared to the regiolects (i.e. for each group, the mean D score is significantly greater than zero. For Antwerp participants in experiment AB: M = 0.24, SE = 0.04, $t(44) = 5.88$, $p < .0001$; for West-Flemish participants in experiment AB: M = 0.36, SE = 0.05, $t(44) = 7.43$, $p < .0001$; for Antwerp participants in experiment CD: M = 0.39, SE = 0.04, $t(44) = 10.30$, $p < .0001$; for West-Flemish participants in experiment CS: M = 0.14, SE = 0.05, $t(40) = 2.92$, $p < .0001$).

predictor	estimate	*p*	
intercept (grand mean)	0.281	< .0001	***
origin participant			
Antwerp	0.035	.11	n.s.
varieties in experiment			
Antwerp-SBD	0.017	.43	n.s.
origin x variety			
Antwerp x Antwerp-SBD	-0.092	< .0001	***

significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 'n.s.' 1

Table 6 - Summary of the linear regression model of the D scores (Chapter 2)

In this study, we are mainly interested in whether there are any differences in attitudes between the Antwerp and West-Flemish participants towards the three varieties presented in our experiments. In other words, our focus lies on the interaction term in the regression analysis (origin x variety), which shows a significant effect (see Table 6). If we tease apart the interaction using post hoc tests (*t*-tests with Bonferroni correction), we find that Antwerp participants show a stronger preference for SBD (M = 0.39, SE = 0.04) compared to their West-Flemish counterparts (M = 0.14, SE = 0.05) in the CD

experiment which contained the standard variety and West-Flemish regiolect, $t(78.83) = 4.21$, $p < .0001$. The difference in D scores between the two participant groups in experiment AB does not reach significance. When the attitudes of the participant groups are compared across experiments, they show a significant pattern of smaller D scores in the experiment containing respondents' own regiolect compared to the experiment containing the other group's regiolect for both Antwerpians (experiment AB: $M = 0.24$, $SE = 0.04$; experiment CD: $M = 0.39$, $SE = 0.04$; $t\ (87.52) = -2.70$, $p < .05$) and West-Flemings (experiment AB: $M = 0.36$, $SE = 0.05$; experiment CD: $M = 0.14$, $SE = 0.05$; $t\ (83.94) = -3.27$, $p < .01$). This pattern can be described as a decrease in participants' preference for SBD when presented with their own regiolect and hence be interpreted as an indication of in-group preference. However, the pattern can just as well be characterized as an increase in preference for the standard variety when presented alongside another group's regiolect which is then perceived as dialectal and triggers a normative reflex. This shows how the results of a P-IAT are essentially contextualised by the specific comparisons of target concepts included in the experiment (see also Chapter 3 on context in the P-IAT).

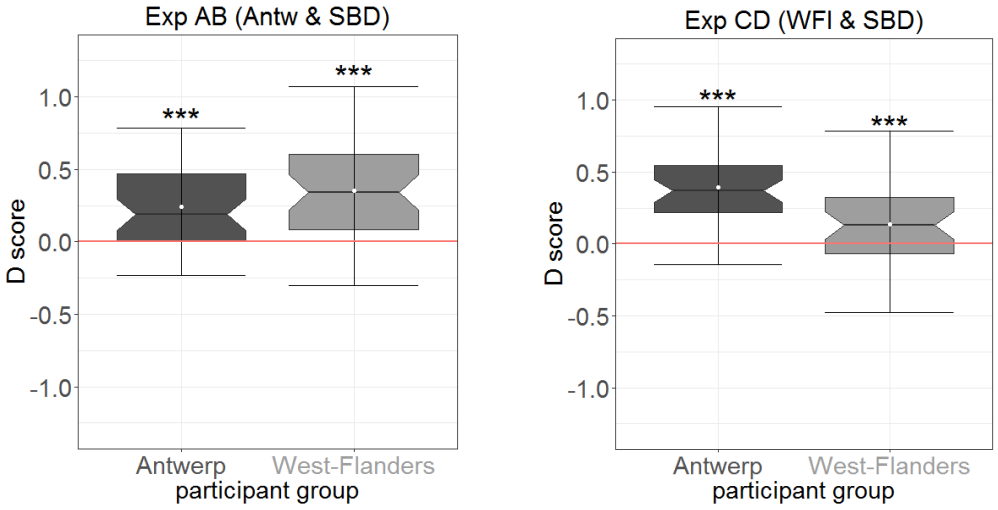


Figure 4 - Boxplots of the D scores in experiment AB and experiment CD (Chapter 2). Positive D scores indicate a preference for the standard, negative D scores a preference for the regional variety included in the experiment.

To summarise, the most prominent finding to come out of the implicit attitude measurement is the overall appreciation for the standard variety. In addition, there was a significant pattern of in-group preference (which was also found in the AP study from Speelman et al., 2013). However, we did not find any evidence of Antwerp participants preferring their own variety over SBD as was the case in Speelman et al. (2013). Potential explanations for this partial divergence between both studies are explored below, in Section 4.1.

3.2 Explicit attitude measurement

The results of the explicit attitude measurement are summarised in Figure 5 and show a similar pattern as in the implicit measurement. Note that due to a technical problem, the explicit attitudes of participants taking experiment D were not recorded correctly. Hence, the analysis for attitudes towards West-Flemish vs. SBD will solely be based on the data collected in experiment C. Because the D scores used to analyse the implicit attitudes in the P-IAT are a relative measure, a difference score was computed between the explicit rating of SBD and the regional variety, in order to make both measures as comparable as possible.

As for the implicit measurement, the outcome of the explicit measurement was modelled using linear regression (with sum coding). The summary of the regression model shows that the grand mean is significantly larger than 0 which is an indication of a general preference for the standard variety (see Table 7). However, if we break up this grand mean and test whether the means per group in each of the experiments show a significant preference for SBD (using *t*-tests with Bonferroni correction), we see that this is only the case for West-Flemish participants in experiment AB (M = 2.71, SE = 0.46, $t(44) = 286$, $p < .0001$) and Antwerp participants in experiment C (M = 4.06, SE = 0.62, $t(17) = 286$, $p < .0001$). In other words, we only see explicit attitudes favouring SBD in participants who were presented with the other group's variety, which could be interpreted as in-group preference or a normative reflex when presented with the other group's regiolect, just like we observed for the P-IAT results.

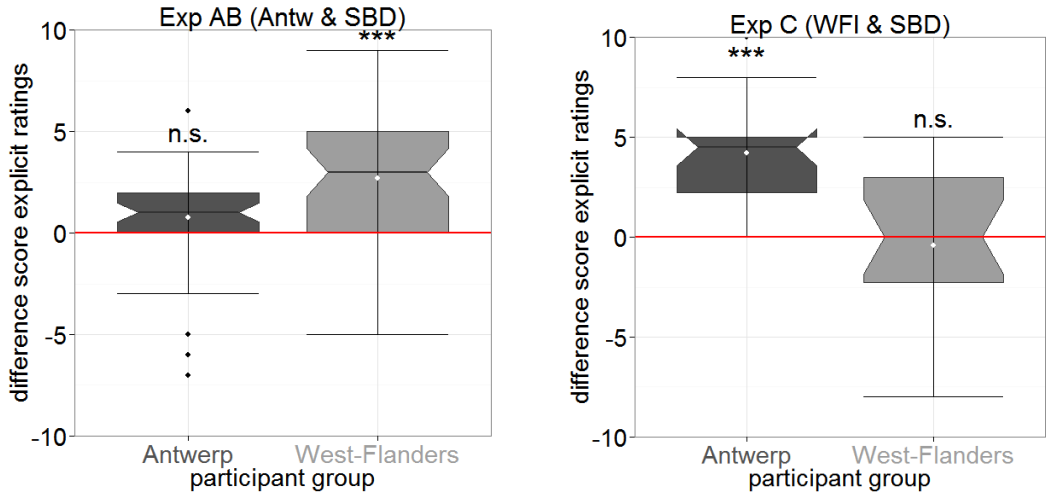


Figure 5 - Boxplots displaying the difference scores between the explicit ratings for SBD and the regional variety in experiments AB and C (Chapter 2). Positive scores indicate a preference for the standard, negative scores a preference for the regional variety included in the experiment.

In the regression model presented in Table 7, the interaction between variety and participant origin is highly significant[22]. A closer inspection of the interaction effect using post-hoc tests (*t*-tests with Bonferroni correction) shows a pattern similar to the P-IAT results. The differences between the two participant groups in both experiment AB (Antwerpians: $M = 0.78$, $SE = 0.39$; West-Flemings: $M = 2.71$, $SE = 0.46$, $t(85.51) = -3.20$, $p < .01$) and experiment C (Antwerpians: $M = 4.23$, $SE = 0.51$; West-Flemings: $M = -0.4$, $SE = 0.84$, $t(31.74) = 4.68$, $p < .001$) are significant. Comparing groups across the experiments, we find that participants from either region show a significantly weaker preference for SBD when presented with their own regiolect, or alternatively, a stronger preference for the standard when confronted with the other group's variety (Antwerp participants: $M_{AB} = 0.78$, $SE_{AB} = 0.39$; $M_C = 4.23$, $SE_C = 0.51$, $t(44.96) = -5.35$, $p < .0001$; West-Flemish participants: $M_{AB} = 2.71$, $SE_{AB} = 0.46$; $M_C = -0.4$, $SE_C = 0.84$, $t(30.91) = 3.23$, $p = .01$). This

[22] 'Origin participant' also shows a significant main effect. However, the fact that 'origin participant' interacts significantly with 'variety' makes the main effect of the former less informative (Baayen 2008: 166). Hence, we do not consider this main effect further. An ANOVA of the model also confirms that 'origin participant' does not significantly contribute to the model.

pattern can again be interpreted as either evidence for in-group preference or as a normative reflex when presented with the other group's regiolect.

predictor	estimate	*p*	
intercept (grand mean)	1.83	< .0001	***
origin participant			
Antwerp	0.67	.02	*
varieties in experiment			
Antwerp-SBD	-0.08	.76	n.s.
origin x variety			
Antwerp x Antwerp-SBD	-1.64	< .0001	***

significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 'n.s.' 1

Table 7 - Summary of the linear regression model of explicit attitude ratings (Chapter 2)

3.3 Correlation analysis implicit-explicit attitude measurement

Spearman's rho was used to compute the correlation between the D scores obtained in the P-IAT experiment and the difference scores collected through the direct ratings of the varieties. Implicit and explicit attitude measurements show a moderate correlation for participant groups which were presented with their own variety (Antwerpians in experiment AB and West-Flemings in experiment C, see Table 8). In both cases the correlation just exceeds conventional significance levels of $p = .05$. In conditions where participants were presented with the other group's variety compared to SBD, results from the implicit and explicit measurements were not correlated.

experiment	participant group	r_s	*p*	
AB	Antwerp	.275	.067	.
	West-Flanders	-.017	n.s.	
C	Antwerp	-.091	n.s.	
	West-Flanders	.424	.062	.

significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 'n.s.' 1

Table 8 - Correlations between implicit and explicit attitude measures (Chapter 2)

4 Discussion

In this section, we will examine the results reported above in more detail (Sections 4.1 and 4.2), as well as take a step back and review the potential of the P-IAT as a measure for language attitudes (Section 4.3). The discussion of the results will be approached in two steps. First, in Section 4.1, we compare the results of both implicit measures, the P-IAT and AP, and discuss why it is that these results do not fully overlap. Secondly, in Section 4.2, we focus on the comparison between the implicit and explicit measurements and discuss the correlation between the two. For this last step, we are restricted to the data collected in our own study, so we can only directly compare the P-IAT outcome with the explicit ratings.

4.1 The implicit attitude measurements: P-IAT vs. AP?

When we compare the results of our implicit attitude measurement using the P-IAT to the AP results reported in Speelman et al. (2013), we find that they mostly converge: in both studies participants are more positive towards their own regional variety than that of the other group, and West-Flemish participants prefer SBD over their own regiolect. However, results diverge on one point: we found a general preference for the standard variety over and above any regional varieties in *both* participant groups, while Speelman and colleagues report Antwerp participants to be slightly more positive towards their own regiolect than towards SBD. No evidence of the latter pattern was found in our data. We discuss three methodological issues that may contribute to this difference, two of them relating to the structure of the P-IAT and one of a more general nature.

Firstly, despite the use of highly similar participant groups and identical stimuli in both studies, the fact remains that they rely on fundamentally different methods, which may explain the partially diverging results. It has been shown that techniques from the IAT paradigm (such as the P-IAT) and priming methods measure different types of constructs (Gast & Rothermund 2010). While AP measures attitudes towards stimuli (in Speelman et al., 2013: audio samples representing actual language use), the IAT measures attitudes

towards both the stimuli in the experiment (in our study: the same audio samples) *and* the labels representing the target and attribute concepts (in our study: 'Antwerp accent', 'West-Flemish accent' and 'neutral accent'). Previous linguistic attitude research has already shown that measuring attitudes towards actual language use and conceptual language varieties represented by a label may not yield the same results (e.g. Bishop et al. 2005a; Kristiansen, 2010c; Grondelaers & Kristiansen 2013). Coupland & Bishop (2007: 85) suggest that attitudes towards labelled varieties are 'broad language-ideological structures', but that these can interact with many factors in contexts of actual language use, potentially resulting in different attitudes.

Secondly, and related to the above, there is another structural aspect of the (P-)IAT that may explain why our results do not exactly match the ones reported in Speelman et al. (2013), namely the measure's comparative nature. Both AP and the P-IAT measure associations with a certain category in comparison to another category. However, in the P-IAT this is more perceptible because of the labels present in the top corners of the screen throughout the experiment. Hence, it is not unlikely that the continuous presence of the category labels in the (P-)IAT make this comparative character more salient. Considering this explicit comparison of varieties against the background of the normative nature of SBD that resulted from the top-down standardisation history of Dutch in Flanders and led to its celebration as the only best language variety (Geeraerts & Van de Velde, 2013), it is possible that the presence of SBD as an explicit category in both experiments AB and CD prevented measuring any positive attitudes towards the regiolects (even though an attempt was made to use a less normative label for SBD). This issue has also been raised by Grondelaers et al. (2011) and Grondelaers & Speelman (2013) in the context of the speaker evaluation paradigm. In that respect, it would have been interesting to compare the results of experiments AB and CD to those of a P-IAT comparing only the regiolects, leaving out SBD. This is something we would like to take up in future research.

Finally, there is always the possibility that the partial divergence of results is not (exclusively) due to the structural nature of the P-IAT compared to AP. There may be hidden variability on the side of the participants that we

have no means of controlling for. Additionally, there is a gap of approximately five years between both studies. Even though we would not expect attitudes towards the varieties under study to have shifted dramatically in this time frame, it is another variable that we cannot control for.

4.2 Correlations between the implicit (P-IAT) and explicit attitude measurements

Overall, the implicit and explicit measurements in this study lead to very similar results. However, we found only moderate correlations or no correlation at all between our implicit (P-IAT) and explicit attitude measurements. Weak correlations between IAT based measures and explicit attitude measures are frequently found in the social psychological literature. Usually correlations between .24 and .37 are reported (Hofmann et al. 2005a; Nosek et al. 2005; Teige-Mocigemba et al. 2010). There are several ways to interpret or explain our modest correlations: there could be a number of methodological issues, in addition to an explanation relating to the degree of social sensitivity of the domain under study.

Firstly, the order in which the implicit and explicit measures are presented may influence the strength of the correlation between their outcomes. Bosson and collaborators (2000) report a stronger correlation if the explicit measure precedes the implicit one. In our study, the measures were presented in the opposite order, which may explain why correlations were moderate at best. However, other studies have failed to find such order effects, yet suggest these effects may occur in case of new, unstable or ambivalent associations, which may or may not apply to the associations measured in our study (Hofmann et al. 2005a; Hofmann et al. 2005b; Nosek et al. 2005). Not only the order of the tasks can influence correlations, approaches to deal with the IAT's block order effects may play a role as well. Counterbalancing block order to control for block order effects, as was done in this study, can introduce additional error variance in the results which may diminish correlation with explicit measures (De Houwer et al. 2015).

A second and perhaps more important methodological issue complicating the comparison of the P-IAT results and the explicit rating task is

'structural fit' (Payne et al. 2008). This term refers to whether or not two methods measure the same type of construct. As pointed out in Section 4.1, the P-IAT measures attitudes towards both the stimuli and labels used in the experiment, while in our explicit rating task, only labels were evaluated. Hence, the fit between both is not ideal. In addition to structural fit, there may be an issue with the conceptual similarity (Hofmann et al. 2005a: 1380) of both our measurements: as discussed earlier, the P-IAT is a relative attitude measure comparing two attitude objects. Our explicit rating task, on the contrary, required absolute evaluations of the varieties.

A third methodological aspect relating to the correlation between our implicit and explicit attitude measurement concerns the phrasing of the question in the explicit rating task: how meaningful was that question for the participants? They were asked how good or bad they thought each of the varieties in the experiment was. Even though none of the participants protested when presented with the rating task or commented on it when given the opportunity after the experiment, it is not inconceivable that this question was not meaningful for them or might have been interpreted in various ways by different participants, leading to small or no correlations between the implicit and explicit measurements.

Another explanation for the moderate or lacking correlations between the implicit and explicit measurements may reside in the degree of social sensitivity of the domain under study. Greenwald et al. (2009) have reported differences in implicit-explicit correlations dependent on the domain of research, which they linked to the degree of social sensitivity of those domains. For instance, they report lower correlations for studies in the domain of racial prejudice (Black vs. White), which is much more socially sensitive and hence, can lead to impression management on the side of the participants, compared to a domain like consumer preferences, where social sensitivity is much less at play. Unfortunately, not much is known about correlations between IAT results and explicit measures for language attitude studies[23]. Besides, social sensitivity

[23] To the best of our knowledge, the only information on implicit-explicit attitude correlations for IATs focusing on language attitudes can be found in Campbell-Kibler (2012), Pantos & Perkins (2012) and McKenzie (2017). Campbell-Kibler (2012: 760)

of attitudes towards certain language phenomena will be highly dependent on the speech community under study. Yet, social sensitivity could potentially explain why we found moderate correlations for participants rating their own varieties, but no correlation whatsoever when measuring their attitudes towards another groups' accent: evaluating the in-group may be less socially sensitive and require less impression management than judging an out-group. This hypothesis would need to be investigated further in future research though.

4.3 The P-IAT as a new measure for language attitudes?

To conclude this discussion we may, with due caution, evaluate the P-IAT as a measure for language attitudes. Based on what is known about the method so far and its use in the current study, what can we conclude about its usefulness as an addition to the traditional array of methods used in linguistic attitude research? In this discussion, we will revisit some of the criteria put forward in the comparative review in Chapter 1 and consider them in more detail.

First and foremost, previous social psychological research has shown that the (P-)IAT is a reliable and valid measure of implicit attitudes (Nosek et al. 2005; Gawronski et al. 2011). It is also difficult to 'fake' an IAT (although not completely impossible; Steffens 2004; Fiedler & Bluemke 2005; Cvencek et al. 2010), which makes it an interesting option to study associations that participants are unwilling to share explicitly or are not aware of. Additionally, in this study, we have been able to use the P-IAT successfully with language stimuli. This gives the method reasonably positive prospects as a new measure for sociolinguists. However, as has already been touched upon in Chapter 1, there are a number of limitations to the P-IAT, some of them shared with the traditional IAT, and certain aspects of the method need further investigation. Issues to be discussed in the following pages include the comparative structure

---

only mentions 'little correlation' between her D scores and explicit measure. Pantos & Perkins (2012: 12) report a small, but significant correlation (r = .26, p = .03). The correlations reported in McKenzie (2017) are small as well (r = .17, p = .086; r = .43, p = .077).

of the P-IAT, practical restrictions on large-scale P-IAT experiments, the selection of suitable stimuli, the need for further research on the categorisation mechanisms at play in the IAT paradigm, and the importance of the notion of structural fit for attitude research in general.

From both a practical and theoretical point of view, it is important to be fully aware of the P-IAT's inherently comparative structure. First of all, the IAT only offers relative attitude measurements without reference to a neutral benchmark. Secondly, the method requires binary target and attribute concepts, which can be inconvenient when, for instance, one wants to study attitudes towards a single language variety without comparison to other varieties. There are alternative methods in social psychology such as the Single Target IAT (Wigboldus et al. 2004) and Single Category IAT (Karpinski & Steinman 2006) which allow non-binary target categories. However, as discussed in Chapter 1, these are incompatible with the use of auditory target stimuli without running into the problem of recoding based on the modality of the stimuli[24] (Gawronski et al. 2011). To avoid this problem, these measures require target and attribute stimuli of the same modality. Yet, if both target and attribute stimuli are presented in auditory form, it is not clear to participants whether they need to be categorised as targets or attributes. To make that clear, there would have to be a difference between both types of spoken stimuli. However, that would create a confound in the experiment. For instance, if you want to measure attitudes towards a single regional variety, you would have to present the spoken attribute stimuli in a different accent. But this second accent would evoke associations of its own, since there is no such thing as an attitudinally neutral language variety.

Despite the (P-)IAT's comparative nature being framed above as potentially inconvenient, we do not believe that it necessarily is a bad characteristic, as it may well be a more ecologically valid way of measuring attitudes than using absolute measures. Judgments about language varieties/variants would seem to be intrinsically relative anyway: when an

[24] I.e. categorising stimuli based on their modality (e.g. picture vs. sound) rather than on the target and attribute categories.

individual judges a certain variety or variant, it will always be against the background of other varieties/variants that individual is familiar with. For example, one may think badly of one's own regional variety compared to the standard variety, yet in comparison to another regional variety, one's own variety may be perceived quite positively. Similarly, language users may have positive associations with a certain variety in context A or used for function X, but not in context B or used for function Y. Although we have not controlled for that type of contextual factors in the present study, they should certainly be explored in future research (for a first exploration with the P-IAT, see Chapter 3). The advantage of the P-IAT's comparative structure then is that it forces the researcher to make explicit the comparative nature of attitudes which lurks in the background in absolute measurements. When using an absolute measure, participants may well be evaluating a variety compared to another variety, but the researcher has no way of controlling what participants are implicitly comparing that variety against. From this perspective, the P-IAT allows us to get a better grip on the contextual nature of language attitudes.

Fully exploiting the comparative nature of the P-IAT, however, means dealing with certain practical restrictions. If one desires to study more than two languages, varieties or variants, the binary structure of the P-IAT will lead to a multiplication of the number of comparisons and hence the number of experiments to be conducted. This entails the added complication of practice and fatigue effects. That is why it is not recommended to have one participant complete multiple consequent experiments, as there is a risk of the P-IAT effect diminishing or disappearing as a result of these practice and fatigue effects in the second and subsequent tests (Gawronski et al. 2011). This means that the number of participants needed for an experiment measuring attitudes towards more than two languages, varieties or variants quickly adds up. These reaction time based tests are traditionally conducted in laboratory settings where participants take the experiment individually in a quiet room in order to avoid any distraction, which means that the use of these measures rapidly becomes highly time-consuming and unattractive for large-scale studies. Yet, previous work in social psychology has shown that it is possible to take the IAT paradigm out of the laboratory and conduct the experiments online (e.g. Friese et al.

2007; Xu et al. 2014). Admittedly, the uncontrolled conditions of online (P-)IATs will entail a number of additional difficulties such as potential distraction due to external, environmental features. But for certain studies, these drawbacks may be outweighed by the advantages, like the potential to reach a larger and more diverse sample of participants and the relative ease to conduct large-scale studies (Nosek et al. 2002a; ibid. 2002b).

Another potentially problematic aspect of the P-IAT, both from a practical and a theoretical point of view, involves the stimuli used in the measure. Despite the relative freedom to use any modality of stimuli one desires, it is hard to select suitable stimuli. From a practical perspective, stimuli need to fulfil many requirements: in addition to being good exemplars of the language phenomenon under study, they have to be controlled for several aspects (e.g. valence or any other feature that may create a confound with the target or attribute categories or can be used for recoding strategies). Fortunately, for several languages norm data are available for concepts like valence (see for instance Moors et al. 2013 for Dutch). Additionally, it is important that IAT stimuli are very short. The longer the stimuli, the more likely it is that the implicit character of the test will be diminished: participants may respond in a less automatic way, if they get more time to process the stimuli. The necessity to use short stimuli also makes it challenging to use the IAT paradigm to study phenomena like syntactic variation which may require longer stimuli.

The length restriction also entails a theoretical issue: the language stimuli in a (P-)IAT are completely decontextualised. Hence, one could question the validity of a language attitude measure if the language presented in the experiment is decontextualised to such a high degree. However, if a memory component of attitudes is assumed – as many psychologists (Albarracín et al. 2008), as well as linguists do (Preston 2015), one could argue that what is being measured here is a type of association that functions as a starting point for, or that feeds into the formation of an evaluation of an attitude object in a certain context. Depending on the context in which the attitude object is encountered, the associations measured with the P-IAT can enter into competition with other information present in that context or in memory, and

may or may not play a role in the formation of a final evaluation (see for instance Campbell-Kibler 2009; ibid. 2012: 761-762 for a similar point of view). In that respect, the associations measured with the P-IAT can provide valuable information for sociolinguistic research despite the high degree of decontextualisation. Yet, it would be interesting and make the P-IAT even more appealing as a method for sociolinguistic research, if contextual factors could somehow be incorporated into the experiment. The interaction between the social meaning of language phenomena and certain types of contexts of use could be studied more systematically. The research available in social psychology on this topic seems encouraging (e.g. Gschwendner et al. 2008 for racial attitudes and the IAT as an anxiety measure). In Chapter 3, we report an experiment that explores the possibility of including situational context in the P-IAT as a language attitude measure.

An aspect of the (P-)IAT that is not entirely understood yet, is the influence of the category labels used in the test. As indicated above, it has been suggested that the (standard) IAT measures associations towards a combination of stimuli and labels (Gast & Rothermund 2010). However, little is known about the categorisation processes at work during the P-IAT. A crucial question in this respect is whether the P-IAT measures associations with categories as represented in participants' mind or whether it measures attitudes towards ad hoc constructed categories imposed by the labels used in the experiment. This is a topic worth exploring further, if we want to be able to get a better grip on how the IAT works and how/whether it can help us to understand how the social meaning of language variation is processed and represented in the brain.

A final issue, we would like to come back to is structural fit, which was introduced in Section 4.2. We consider structural fit of crucial importance to attitude research. It is vital in order to understand what each attitude measure is most suitable for and how its results compare to other measurements. If we put the measures considered in this study on a continuum based on what type of construct they measure, we get a picture that matches the trends observed in the results from those respective measures quite nicely. On one extreme of the continuum, we could place AP which measures attitudes towards a collection

of stimuli. Somewhere in the middle we find the P-IAT in which these stimuli play a role as well, but the category labels are a substantial part of the construct that is being measured too. Our explicit rating task would then be the other extreme of the continuum focusing exclusively on labels. However, maybe we should review the position of the rating task slightly, given that participants had just been presented with multiple stimuli during the preceding P-IAT. This means that the structural fit between the P-IAT and explicit rating task is perhaps slightly better than that between the P-IAT and AP. This observation seems to be reflected in the results: those obtained with the P-IAT and explicit ratings show more similar trends than those from the AP experiment. We believe all attitude research, whether it uses recently developed social psychological measures or more traditional sociolinguistic methods, should consider structural fit carefully when choosing the appropriate methodology for its purposes or when comparing results from different measures.

5 Conclusion

In conclusion, it seems fair to say that the P-IAT (and the IAT paradigm in general) has considerable potential as a measure for language attitudes. Like any method in the field, the P-IAT comes with a number of intrinsic limitations and certain aspects of the method are not yet fully understood. Further exploration of the P-IAT's possibilities and characteristics is certainly required. Yet, with due caution pending further research, we venture that the P-IAT is a promising new method to add to the (socio)linguist's toolbox. In no way do we mean to suggest that this social psychological method could replace the existing array of methods at the disposition of the language attitudes researcher, but we firmly believe it can provide interesting insights, when used with due consideration of its limitations. As Garrett (2005: 1257-1258; 2010: 201-222) indicates, the best insights into a language attitude landscape can be obtained by combining a diverse range of methods. We have presented evidence that the P-IAT can be one of those in future language attitude research.

Chapter 3

Exploring the potential of the Personalized Implicit Association Test to measure language attitudes in context

1 Introduction

Social psychology has recognised and demonstrated that contextual factors influence attitudes (Albarracín et al. 2008, Ferguson & Bargh 2007; Gawronski & De Houwer 2014; Wittenbrink & Schwarz 2007). In a similar vein, sociolinguistics has investigated the influence of diverse contexts on attitudes towards language (variation). Several studies have tried to explore how the social meaning of language variation is moderated by different types of contextual information (e.g. Campbell-Kibler 2010, Campbell-Kibler & McCullough 2015). Yet, despite their evidence that context plays a crucial role in moderating (language) attitudes, context features are often still ignored in quantitative language attitude research (Soukup 2013b).

Context dependence of language attitudes is a particularly relevant topic in the light of the recent introduction of social psychological implicit attitude

measures in sociolinguistic research, like affective priming (Speelman et al. 2013) and the Implicit Association Test (IAT, Pantos 2012, Rosseel et al. under review). An aspect shared by many of these reaction time based attitude measures is that they present the attitude object void of context features. This is usually the result of restrictions placed on the stimuli used in these tasks: they have to be as short as possible in order to assure that the method measures the outcome of automatic processes. The required shortness of the stimuli in these methods makes it difficult to study attitudes towards linguistic phenomena that cannot easily be captured in a short word (e.g. syntactic structures or certain intonation patterns). But it also makes it difficult to embed the linguistic variant or variety under study in a wider context, linguistic or otherwise. Hence, the study reported in this chapter sets out to explore the possibilities of incorporating contextual features in the experimental design of one such social psychological measure, the Personalized Implicit Association Test (P-IAT). More specifically, we attempted to manipulate situational context while using the P-IAT. Despite a number of social psychological studies reporting the successful introduction of context features within the IAT paradigm, this study comes to the conclusion that including context in a linguistic version of the P-IAT is not straightforward and further research or methodological improvement is warranted if sociolinguists intend to begin using the measure to study the influence of context on language attitudes.

Before describing the design of our study (Section 2) and reporting the results (Section 3), we will briefly discuss previous sociolinguistic work on the context dependence of the social meaning of language variation (Section 1.1). This is followed by a short overview of the ways in which social psychologists have attempted to incorporate context into the design of the P-IAT (Section 1.2). As a final part of this section, some background will be provided about the language varieties and speech community that were part of the study (Section 1.3). Combined with the reviews in Section 1.1 and 1.2, this will allow us to frame the research questions and formulate hypotheses for the study.

1.1 Context and the social meaning of language variation

In contemporary sociolinguistics, language variation is considered as one of many tools people have at their disposal to convey social meaning, or as Eckert (2012: 91) puts it: language variation is a 'component of a broader semiotic system.' Multiple social meanings can be associated with a linguistic variable, forming an indexical field of related meanings. Yet, which of those potential meanings is activated in interaction depends on the context in which the variable is used (Eckert 2008). Consequently, taking context into account should be a primary concern when studying language attitudes.

Many linguistic studies have provided evidence for the pivotal role of context in relation to the social meaning of language variation. When inspecting these studies, it becomes apparent that we can interpret context in its broadest sense, ranging from linguistic to situational context. Without attempting to provide an exhaustive typology of different types of context[25], let us illustrate the breath of context dependency in language attitudes by giving a few examples. Firstly, it is well documented that various types of speaker information may affect attitudes towards a linguistic feature. Aspects like perceived social class, regional origin, profession or gender of a speaker can impact attitudes towards that speaker's language (e.g. Campbell-Kibler 2007; ibid. 2009; ibid. 2010; Nelson et al. 2016). In a recent study, Campbell-Kibler and McCullough (2015) also demonstrated the relationship between physical features of a speaker and attitudes towards their speech: perceived accentedness was shown to influence how well participants thought a certain voice and face matched. A second type of context that plays a crucial role for the social meaning of language features is their linguistic context. The (bundles of) other linguistic features a variant co-occurs with may influence its social meaning (e.g. Levon 2007 on pitch range and sibilant duration in the perception of gender and sexual identity; Pharao et al. 2014 on the social meanings of /s/ in Danish depending on whether a variant co-occurs with

[25] See for instance Gallois & Callan (1985) for a proposition of a typology of different types of contexts in the light of their interaction with language attitudes and speech perception.

features typical of Modern Copenhagen speech or typical of street language). Other varieties present during an interaction or varieties a hearer was previously exposed to can co-determine how a linguistic feature or variety is evaluated as well (e.g. Price et al. 1983; Abrams & Hogg 1987; Walker et al. 2014). A final type of context we will mention here and the type that concerns the study described below is situational context (e.g. Giles & Ryan 1982; Creber & Giles 1983; Gallois & Callan 1987; Cargile et al. 1994, Cargile 1997; Soukup 2013a; ibid. 2013b). Situational context has many facets, but one that is often highlighted in relation to language attitudes is the degree of formality of a situation. Street and colleagues (1984), for instance, report that speakers are rated differently for competence in informal conversation compared to a more formal job interview. Similarly, Creber & Giles (1983) demonstrate that a formal (school) vs. informal (youth club) situational context influences status evaluations of a regional variety of English and received pronunciation (RP). Note that the different types of context mentioned above co-occur in interaction. As a result, trying to tease them apart and studying the influence on attitudes for each type of context separately will always be difficult, and to a certain extent an non-ecological undertaking.

Context does not only affect the activation of social meaning of language variation, it also influences language perception and processing. Information about a speaker's (social) background can influence what listeners think they hear. In her classic 1999 study, Niedzielski showed that the imagined regional background of a speaker determined how the respondents in her experiment classified tokens of a linguistic variable that was stereotypically associated with one region, but not with the other, even though it occurred in both. When participants were told that the speaker was a Canadian, whose speech is stereotypically associated with a raised pronunciation of /aw/, they identified raised tokens as such. However, when listeners thought they heard a fellow Detroiter, they failed to perceive the same tokens as raised. Later studies have reported similar effects of many kinds of speaker information on speech perception ranging from speaker's age (e.g. Hay et al. 2006b; Koops et al. 2008; Drager 2011) to social class (e.g. Hay et al. 2006b), gender (e.g. Strand 1999, Johnson et al. 1999) and ethnicity (e.g. Rubin 1992; Staum Casasanto 2008).

In addition to the influence of speaker information, other contextual cues, such as previously heard language varieties, have also been shown to influence speech perception (Hay et al. 2010).

Despite the plentiful evidence available on the context dependency of language attitudes, quantitative experimental language attitude research is criticized – mainly from a qualitative, constructionist point of view – for disregarding the crucial role of interactional context (Soukup 2013b). Many studies present participants with stimuli that are devoid of any context: either stimuli are so short that linguistic context is limited, or no information on the speaker or situational context are provided, or both. One explanation for this is that experimental designs are often restrictive and make the inclusion of contextual information challenging. Alternatively, including contextual elements may introduce confounds or additional variation in the outcome of the experiment that is difficult to control. Yet if one aims to study the social meaning of language variation in a more ecologically valid way[26], attempts should be made to include contextual features into the experimental design (see for instance Pantos & Perkins 2012 for their explicit attitude measurement; Soukup 2013b; Vandekerckhove & Cuvelier 2007). An excellent example of including situational context into the experimental design of a quantitative language attitude study is the 'newscast paradigm' introduced by Labov et al. (2011). In these studies, participants are presented with a number of audio clips. They are told these clips are recorded by someone applying for a job as a newscaster as part of their portfolio. An experimental setup like this one allows to evoke a specific situational context, which enables the researcher to tap into specific domains of the indexical field of the language features or varieties under study.

From a theoretical point of view, how can contextual influence on the social meaning (and perception) of language be explained? To offer an answer

---

[26] Of course the experimental setting itself is a contextual factor that has its impact on the attitudes measured as well. In that sense, the ecological validity of an experiment will always be limited. Hence, as mentioned in the conclusion of Chapter 2, the best understanding of the social meaning of a language variety or variant is obtained by combining different methodological approaches ranging from quantitative experiments to interactional discourse studies.

to this question, we need an account of how language usage gets stored mentally including the social and contextual information that was available during usage situations. One such account is exemplar theory (e.g. Pierrehumbert 2001, Hay et al. 2006a, Foulkes & Docherty 2006). Taking a broader perspective, exemplar theory links up with prototypicality, which is a theory of categorisation that has played an important role in the field of Cognitive Linguistics. In the analysis of prototype theory presented by Geeraerts (2010b), exemplar theory corresponds to an extensional perspective on prototypicality. What makes a prototype view of categories different from a traditional perspective is that the former recognises that not all members (i.e. exemplars) of a category have to be equal and that categories are not necessarily discrete entities, but can have fuzzy edges. This means that some exemplars may be more salient than others and that it may be unclear for certain exemplars whether they are a member of the category or not. Before returning to the relevance of exemplar theory for context and language attitudes, we refer to Chapter 4 and Chapter 5 for further discussions of the importance of categorisation processes for language attitudes and the potential of prototypicality as a component of Cognitive (Socio)linguistics to account for these issues.

Exemplar models of language are based on the idea that language is stored in the brain in the form of exemplars that are built on past experience. Every time linguistic input is encountered, traces of that input are stored in the brain and more abstract categories are derived from these traces. New input can alter the existing representation and old traces can erode if not frequently activated (Squires 2013). These cognitive representations of language not only store detailed information about the form and denotational meaning of a linguistic token, they also include knowledge about the social context it was encountered in (Foulkes & Docherty 2006). Hence, these representations comprise a network of associations between various social meanings of linguistic features and knowledge about the different types of context they occurred in. A linguistic feature presented in one context may then activate different social meanings compared to when it is presented in another context, depending on how frequently a person has encountered that feature in different

social settings. So for instance, as reported by Creber & Giles (1983), the social meaning of RP in terms of status increased when the variety was presented in a formal context compared to when it was presented in a more informal situation. Conversely, starting from the social context rather than the linguistic information, contextual cues are drawn on by speakers in speech perception as the examples above illustrated. For instance, Staum Casasanto (2008) shows how people use social information about speaker ethnicity they have stored, to resolve ambiguity following from potential d/t deletion in speech perception. To sum up, exemplar theory predicts that the encounter of a linguistic feature may activate social knowledge, and social cues may activate certain linguistic tokens (Squires 2013).

1.2 Incorporating context in the IAT

In the section above, we have illustrated the non-negligible role of context for language attitudes and argued that context should be an integral part of language attitude experiments. Yet, the new methods to study language attitudes that have been imported from social psychology into sociolinguistics make the inclusion of context in experimental designs challenging. As discussed in Chapters 1 and 2, these methods include affective priming (AP, Fazio et al. 1986; Speelman et al. 2013) and the Implicit Association Test paradigm (IAT, Greenwald et al. 1998; Redinger 2010; Pantos & Perkins 2012). Both methods are reaction time based implicit attitude measures (see Chapter 1, Section 2.2). One important aspect of their implicit character is the availability of time (De Houwer et al. 2009, see also Chapter 1, Section 2.2). Participants should have limited time to process the stimuli if the circumstances under which the attitudes are measured are to qualify as automatic. In order to ensure this, stimuli used in AP and the IAT have to be as short as possible. For linguistic stimuli, this means one is limited to short words (written or spoken), as participants need to be able to process them in a matter of milliseconds rather than seconds. As indicated above, this does not only restrict the possibility to measure associations with longer linguistic features (e.g. syntactic structures or discourse-level variables), it also restricts the options for including context in these methods. If there were ways to bring

context into the design of these social psychological attitude measures, that would make them more attractive tools for sociolinguists aiming to study the social meaning of language variation in a more ecological way or for those specifically interested in investigating the impact of certain contextual cues on language attitudes.

Turning to the field of origin of these implicit measures, what is known about the relation between attitudes and context there? Social psychologists do not fully agree on the theoretical models of the cognitive status of attitudes. Some believe attitudes are stored in memory and retrieved when an attitude object is encountered (e.g. Fazio 2007). Others take a constructionist perspective and theorize attitudes as constructed online (e.g. Schwarz 2007). In this view, external contextual input is more important when evaluating an attitude object than information stored in memory (Albarracín et al. 2008). Strong versions of this constructionist model have even argued that no information needs to be retrieved from memory at all and that each time an attitude object is encountered, the evaluation is constructed from scratch (e.g. Schwarz & Bohner 2001). Still other psychologists take a more nuanced position and try to combine the memory-based and the constructionist perspective on the cognitive status of attitudes by recognizing the importance of memory, as well as that of external contextual input and online processing (e.g. Gawronski & Bodenhausen 2006). Despite these different views and the different approaches they entail to explaining the relationship between context and attitudes, psychologist do generally recognise the fact that attitudes are subject to contextual influence (Gawronski & De Houwer 2014). In that light, researchers have explored the context sensitivity of methods like AP and the IAT. Given that the study described in this chapter uses a method from the IAT paradigm, the remainder of this review of context and social psychological implicit attitude measures will focus on the IAT.

A considerable number of studies have shown that it is possible to manipulate context while using the IAT (Ferguson & Bargh 2007; Gawronski & Sritharan 2010). These successful context manipulations can be divided into two categories: (a) experiments where context is evoked before participants take the IAT, (b) experiments where context is brought in during the IAT

procedure itself, either by manipulating the stimuli, or by including context cues in another part of the IAT procedure. The majority of studies that have manipulated context using the IAT fall in the (a) category. Within this group of studies, there is, however, still considerable variation in the way in which context cues are presented. Some studies present participants with visual materials, like pictures (Dasgupta & Greenwald 2001) or video fragments (Wittenbrink et al. 2001). The latter study dealt with prejudice against Black people and asked its participants to watch a video presenting Black people in two different settings. One showed a family gathering, the other depicted a graffitied inner-city scene. Participants were then asked to produce a text about the video they saw, before starting an IAT measuring racial attitudes. Other studies had respondents listen to different types of music (Rudman & Lee 2002), read a text (Foroni & Mayr 2005) or fill out a survey (Steele & Ambady 2006) to evoke particular contexts. Even instructing participants to imagine certain scenarios or qualities before starting the IAT worked to activate contextual cues (e.g. Blair & al. 2001, but see Gawronski & Bodenhausen 2005 for some nuance). Bohner and colleagues (2008), for instance, asked binational individuals to actively think about one of their identities before completing an IAT measuring associations with both their identities. Their results showed that making one of the participants' identities more salient, by having them think about it, influenced the outcome of the IAT (albeit only for male respondents). Finally, characteristics of people present in the experimental setting can also function as contextual cues. Lowery et al. (2001) manipulated the experiment leader's ethnicity in the context of a racial attitude IAT, and reported reduced prejudice against Black people when participants interacted with a Black experimenter compared to a White one. Richeson & Amady (2001) showed that the anticipated role in interaction (in terms of hierarchy) with someone of another ethnicity could function as a contextual cue before completing the IAT. Social role was also evoked in a study by Uhlmann & Swanson (2004), but they employed an aggressive video game to allow participants to assume a specific role.

The second group of studies (b) includes context cues in some part of the IAT procedure itself, rather than presenting them beforehand. Some

studies have successfully manipulated how participants interpret the target category by carefully selecting the target stimuli. Mitchell and collaborators (2003) relied on additional information about the individuals who made up the target categories 'Black' and 'White' in their ethnic prejudice IAT: they used well-liked Black and disliked White persons as representatives of the two categories respectively. Similarly, Govan & Williams (2004) used specific target stimuli to re-define the target categories in their flower/insect, Black/White and plant/animal IATs. Gschwendner et al. (2008), on the other hand, did not manipulate the stimuli in their experiment. Rather, they chose to display a background picture evoking different contexts in an anxiety IAT and a racial attitude IAT. A study combining context manipulation (a) before and (b) during the IAT is Wittenbrink et al. (2001). In addition to their video and writing task in anticipation of the IAT, they also showed snippets of the video participants watched beforehand at random intervals during the IAT.

A final note on context and the IAT relates to the IAT's structure. Requiring a binary target and attribute category (for variations of the measure offering an alternative structure see the discussion in Chapter 1, Section 2.2.5), the IAT already contextualises its target categories in the sense that attitudes to these categories are measured relative to one another. An illustration of this is found in Houben & Wiers (2006). This study used IATs to measure associations with alcohol and found that alcohol associations varied depending on what was chosen as the contrasting second target category (soda vs. animals).

In the study below, we use a personalized version of the IAT, the P-IAT, rather than the traditional IAT (see Chapter 1, Section 2.2.3 and Chapter 2). Although most studies reported above have used traditional IATs or other variants than the P-IAT, there is no reason to suspect that the P-IAT would behave any differently regarding the introduction of contextual cues.

1.3 Background and research questions

In the study reported in this chapter, we set out to measure attitudes towards two varieties of Belgian Dutch in different situational contexts. The varieties under study are standard accented Belgian Dutch (SBD) and a variety of

Colloquial Belgian Dutch (CBD), namely Limburg[27] accented Dutch. The situational contexts chosen for the study are formal vs. informal settings. To understand this choice of varieties and situations, some background on language variation in Dutch speaking Belgium is required. The stratificational structure of Dutch in Belgium anno 2017 can be described as a diaglossic situation, to use Auer's (2005) classification (Geeraerts & Van de Velde 2013; Ghyselen 2016; Geeraerts 2017). There is a continuum with the local base dialects on one end and SBD on the other (see the cone in Figure 6 where the surface of the ellipses represents the amount of variation at one point of the continuum and the distance between the ellipses the linguistic distance between the varieties). CBD is to be situated on the continuum between the two extreme ends. This variety includes features of all linguistic levels which may co-occur in different combinations, and which differ in their degree of colloquiality. This makes it difficult to delineate the variety and allows it to move closer to either end of the continuum (toward the SBD pole in case of fewer and less colloquial features or toward the dialect pole in case of many or highly colloquial/regional features; Geeraerts & Van de Velde 2013; Geeraerts 2017). As explained in Chapter 2, CBD is also characterised by regionally flavoured accents (Geeraerts & Van de Velde 2013). Given that the P-IAT limits the length of the stimuli that can be used, it was decided to use two-syllable words which were recorded in the accents representing the two varieties under study. Hence, the linguistic features included in the stimuli were restricted to phonetic and intonational features.

Previous research on language production has linked the varieties depicted in Figure 6 to a continuum of situational settings and frames reaching from highly informal at the bottom of the cone to strictly formal at the top end (Plevoets 2008; De Caluwe 2009; Geeraerts & Van de Velde 2013; Van Hoof 2013; Van De Mieroop et al. 2016). Additionally, there is some perception research that points in the same direction: Lybaert (2014) found that the participants she interviewed about their perception of SBD, CBD and dialect,

---

[27] Limburg is the easternmost province of Flanders, the Dutch speaking part of Belgium, and is considered part of the linguistic periphery of this Dutch speaking area (Geeraerts et al. 1999; Impe & Speelman 2007)

generally indicated they thought SBD was the variety par excellence for formal situations, particularly for news broadcasts. CBD was associated with informal situations, although there was more variation in participants' responses here. Multiple respondents explicitly dismissed its use in highly formal contexts like news broadcasts on TV or radio. Dialect was only deemed suitable in the most informal and private settings. In a similar vein, Delarue & Lybaert (2016: 254) report that Flemish school teachers say they try to use SBD in formal settings (e.g. giving instructions), but prefer using CBD in more informal interactions with their pupils. Building on exemplar theory (cf. supra), we can explain this link between production and perception: if speakers regularly encounter a variety in a certain situational context, they will store traces of this contextual information alongside other information about the variety and hence will come to associate the variety and situational context. Based on this, we expected participants in our study to hold more positive evaluations about the two varieties under scrutiny when they are presented in their respective situational contexts.

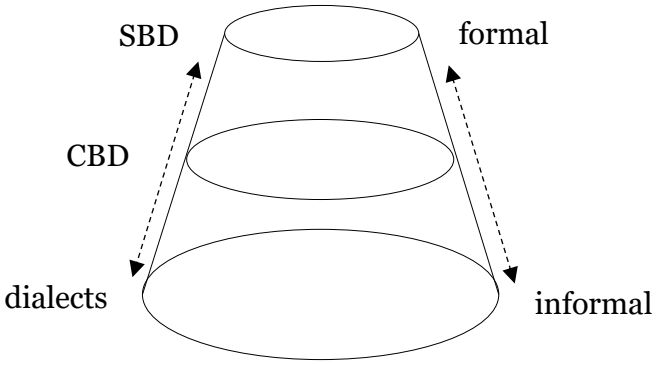


Figure 6 - Stratificational continuum of Dutch in Flanders, based on Geeraerts (2017)

The concrete situations that were chosen to represent formal and informal contexts in our study were TV news broadcasts and socialising (having food or drinks) with friends and family. Note that some linguists (e.g. De Caluwe 2009) believe that CBD is entering domains that used to be the exclusive territory of SBD. Yet, the situational context that all researchers seem

to agree is immune to this supposed invasion is the TV newscast. In Flanders, SBD is strongly associated with the TV and radio news, especially with the news on VRT, the Flemish public broadcasting company, which is notorious for monitoring its newscasters' language (Grondelaers et al. 2016). As mentioned above, evidence of this link was also found in Lybaert (2014). As a result, we decided that TV newscasts would be the situational context that most strongly evokes a formal setting that is related with SBD.

Based on the above, the following research questions were formulated: (1) Can we find evidence that SBD is positively evaluated in formal situations and CBD in informal situations? (2) Is it possible to measure these context-dependent language attitudes using the P-IAT? and (3) What is the most efficient way to implement situational context in the P-IAT? In order to formulate an answer to these questions, a study was carried out which included an implicit measurement of attitudes towards SBD and CBD in a formal vs. informal context using the P-IAT, as well as an explicit attitude measurement involving the same varieties and contexts. This second measurement was included to serve as a point of comparison to assess the results obtained in the P-IAT. In order to get a better understanding of how situational context can best be included in a P-IAT, contextual cues were incorporated in two different ways (for details see Section 2). Given there was no reason to assume a discrepancy between implicit and explicit attitudes, we expected to find positive attitudes towards SBD in the formal context and positive attitudes towards CBD in the informal context. As for the type of context cues, the study was of an explorative nature and no specific hypotheses were formulated as to which way of including the context cues would be the most successful. Note that in order to control for in-/out-group biases (cf. Chapter 2), it was decided to only recruit participants from the Limburg region of Belgium whose regional variety represented CBD in this study.

2 Method

2.1 Participants

In total, 161 participants took part in the study. All were students in higher education recruited at the University of Leuven and spent most of their childhood and teenage life in Limburg, the easternmost province of Flanders. The sample was roughly balanced for gender and age (55% female; M_{age} = 20.96). One participant was removed from the analyses, as they did not complete all tasks in the experiment. Additionally, 9 participants were removed from the sample due to the high number of mistakes they made during the P IAT (i.e. > 20%; Greenwald et al. 2003; Gawronski et al. 2011).

2.2 Design and procedure

Participants first completed two P-IATs which measured their positive/negative associations with Limburg accented speech, their own regional variety of Dutch, compared to SBD. The two P-IATs were manipulated for formality of context: one test contained pictures suggestive of informal settings, while the other evoked formal situations. The order of presentation of the two contexts was controlled between participants. Additionally, we tested two ways in which pictures were used to represent situational context: displaying a background picture throughout the categorisation tasks of the P-IAT (cf. Gschwedner et al. 2008) versus alternating the different blocks of the P-IAT with short sequences of multiple pictures evoking the same situational contexts[28]. The type of context presentation was a between subject manipulation as well. After completing the two P-IATs, participants were asked to fill out a brief questionnaire containing both explicit attitude rating scales and basic demographic questions. This last set of questions allowed us to make

[28] This type of context presentation was inspired by Wittenbrink et al. (2001) who showed snippets of a video participants had watched before the IAT, at random intervals throughout the IAT. In our study, however, participants saw the same sequence of pictures before every block of the IAT. The advantage of this sequence of pictures compared to a single background picture is that it allows a more precise and nuanced image of the situational contexts under study to be evoked.

sure the sample was more or less homogeneous in terms of gender, age and regional background. Finally, respondents were fully debriefed about the study.

The experiments were completed individually in quiet, dimly lit rooms on laptops with 1366 x 768 resolution screens using Affect 4.0 (Spruyt et al. 2010). For the auditory stimuli in the P-IATs, a Jabra UC VOICE 150 MS Duo headset was used. Explicit attitude ratings and the questionnaire collecting demographic information were completed on paper.

The design of the experiment, including participant numbers, is summarised in Table 9. Note that there are eight different conditions (A-H) in the experiment as a result of counterbalancing for three aspects. Two were already mentioned above: context order (under 'context' in Table 9) and type of context presentation ( /for the background picture vs.  /  for the succession of images between blocks of trials in Table 9). A third variable that was controlled for is block order. It is known that the (P-)IAT produces larger effects if the congruent block precedes the incongruent block, hence block order was controlled for between participants (Teige-Mocigemba et al. 2010, see also Chapter 2). The two block orders are presented in Table 9 as BO1 (i.e. response key mapping in the first experimental block: SBD/positive – CBD/negative) and BO2 (i.e. response key mapping in the first experimental block: CBD/positive – SBD/negative). Note that the type of context (formal vs. informal) was a within subject manipulation.

2.3 Materials

The P-IAT requires two types of stimuli: target stimuli which represent the target concept (i.e. language variety) and attribute stimuli representative of the attribute concept (i.e. valence). Hence, we needed target stimuli for both the regional and the standard variety, and positively and negatively valenced attribute stimuli.

condition	P-IAT	context presentation	block order	context	# participants
A	1		BO1	formal	19
	2		BO1	informal	
B	1		BO2	formal	19
	2		BO2	informal	
C	1		BO1	informal	19
	2		BO1	formal	
D	1		BO2	informal	19
	2		BO2	formal	
E	1		BO1	formal	16
	2		BO1	informal	
F	1		BO2	formal	20
	2		BO2	informal	
G	1		BO1	informal	19
	2		BO1	formal	
H	1		BO2	informal	20
	2		BO2	formal	

Table 9 - Experimental design with participant numbers (Chapter 3)

For the target category, we used a set of auditory stimuli. To be suitable, the words we selected as stimuli had to meet a number of criteria regarding length, frequency, semantics, and phonetic structure. First, the P-IAT requires its stimuli to be short in order not to diminish the automatic character of the associations it measures. Hence we limited the length of the target stimuli to two syllables. Second, we avoided words with a low frequency (no words with

under 50 occurrences per million words). Then, regarding semantics, the target words had to be neutral (average ratings between 3.5 and 4.5 on a 7 point rating scale) and not refer to the situational contexts under study, in order to avoid confounds with either the attribute concept or the context factor respectively. For both frequency counts and information on valence, norm data collected by Moors et al. (2013) were consulted. Finally, the stimuli needed to contain phonemes that have a distinctive realisation in the Limburg regiolect compared to SBD, as well as to other regional varieties in Belgium. A list of phonetic variables that meet this criterion was compiled using modern pronunciation manuals that aim to teach their readers how to mask their regional accent (Fraeters & Van Avermaet 2010; Huybrechts et al. 1998; ibid.1999; Timmermans 2008; Van Maele 1984). In total, 203 words with up to three loci for regional pronunciation and satisfying all other criteria were collected.

All 203 words were recorded by five male speakers from different parts of Limburg, who still lived in the area or had spent most of their life there. Each speaker produced both a regional and a standard realisation of every word. For the regiolect guise, they were instructed to speak as if they were talking to someone from Limburg, but not from their own locality, while for the SBD guise they were asked to produce a standard pronunciation aiming for the speech of news anchors on TV or radio. They were not instructed about any specific variables in order to obtain speech samples that were as naturally sounding as possible.

Out of the recordings, an initial selection of words containing maximally regional/standard variable realisations was made by the researcher. This first selection of stimuli was subsequently rated by a sample of five non-linguistically trained listener-judges on how recognisable they were as either Limburg accented or standard accented speech. Stimuli from three speakers were discarded, either because the listener-judges commented on the speaker's quality of voice or because the distinction between their two guises was not clear enough to the untrained ear. Based on those ratings, 6 words, produced by three different speakers, were selected as target stimuli (see Table 10).

target		Attribute		context
label	stimuli	label	stimuli	stimuli
neutrale uitspraak (‘neutral pronunciation’)	*anker*[A, SBD] (‘anchor’) *majoor*[B, SBD] (‘major’) *manier*[C, SBD] (‘manner’) *moment*[B, SBD] (‘moment’) *rechtop*[C, SBD] (‘upright’) *sergeant*[A, SBD] (‘sergeant’)	*vind ik goed* (‘I like’)	5 positive real-life colour pictures	real-life colour pictures representing: - a formal setting (TV newscast) or - an informal setting (chatting with friends/family over dinner/drinks)
Limburgse uitspraak (‘Limburg pronunciation’)	*anker*[A, LIM] (‘anchor’) *majoor*[B, LIM] (‘major’) *manier*[C, LIM] (‘manner’) *moment*[B, LIM] (‘moment’) *rechtop*[C, LIM] (‘upright’) *sergeant*[A, LIM] (‘sergeant’)	*vind ik slecht* (‘I don’t like’)	5 negative real-life colour pictures	

[A] produced by speaker A, [B] produced by speaker B, [C] produced by speaker C

[SBD] realised in the SBD guise, [LIM] realised in the regional guise

Table 10 - Stimulus set for the study in Chapter 3

The set of attribute stimuli consisted of 5 positive and 5 negative real-life colour photographs. They were selected from a set of pictures for which norm data on valence had been collected in previous studies (Spruyt et al. 2002). All pictures measured 410 x 308 pixels.

For the stimuli used to prime situational context, we selected a set of real-life pictures representing either a formal or an informal context. A TV news setting was chosen as the formal situational context, given that in Dutch speaking Belgium, this is the formal situation par excellence in which SBD is spoken and expected (see Section 1.3). In order to avoid unwanted effects due to associations with certain Belgian news anchors or TV channels, we used pictures of foreign news anchors and broadcasting studios which participants were unlikely to be familiar with, but which were highly recognisable as TV news settings for our Belgian participants. All logos and headlines were removed from the stimuli to make them look as generic as possible. As for the informal setting, pictures of friends and family gathering for drinks or dinner were used. For each situational context, 16 pictures were selected. As indicated above, context was primed in two ways in the experiment: for half of the participants the 16 context pictures were displayed between the blocks of the P-IAT, for the other half, one picture was used as a backdrop during all blocks. In the former condition, each picture was displayed for 800 ms with a 100 ms interval between pictures.

The explicit attitude measurement was carried out using two types of 7 point semantic differential scales (see Appendix A). On the one hand, participants were presented with a relative rating task in which they had to indicate which variety they preferred in each of the two contexts (rating scales a. in Appendix A). This rating scale was designed to mimic the relative nature of the P-IAT as closely as possible. On the other hand, we provided two absolute rating scales, so participants could evaluate each variety separately in both contexts (rating scales b. and c. in Appendix A). The motivation for including these additional absolute rating scales was to give participants the opportunity to express for instance positive attitudes towards both varieties, which is impossible in the forced choice task implied in the relative rating scales.

2.4 Analysis

Reaction times recorded in the P-IATs were used to calculate D scores using the D_6 algorithm (Greenwald et al. 2003; Martin 2015). D scores are average difference scores between the experimental blocks that take into account individual differences in speed and add a penalty for errors. In the results below, positive D scores indicate a preference for SBD while negative scores indicate a stronger association between the regional variety and liking. When the D score equals zero, i.e. when participants reacted equally fast in both experimental blocks, there is no association in either direction or the association between target and attribute categories is equally strong.

Note that the experiment was programmed in such a way that the presentation order of stimuli was random, but that no two identical stimuli could follow each other. However, a sequence of two identical stimuli was still possible if one appeared as the final stimulus of experimental block 3 or 6 and the other one as the first stimulus of the consecutive block (respectively block 4 or 7). In the few instances this occurred, the reaction time of the second instance of the stimulus was recoded to be identical to that of its first occurrence in order to avoid undue influence of the repetition on reaction times.

To measure the effect of context, we calculated a difference score between the D scores in participants' formal and informal P-IAT ($E_{context} = D_{for} - D_{inf}$). In other words, we compared participants' relative preference for standard vs. Limburg accented speech in the formal context (D_{for}) with that relative preference for standard vs. Limburg accented speech in the informal context (D_{inf}). As explained above, positive values of D_{for} and D_{inf} indicate a preference for SBD vs. the regional variety in the formal and informal context respectively. Positive $E_{context}$ scores then represent higher D scores in formal contexts compared to informal contexts and indicate behaviour that aligns with what we hypothesize. Negative $E_{context}$ scores indicate the opposite and a score equalling zero means no effect of the context manipulation at all. Positive $E_{context}$ scores can stem from three possible scenarios: (1) the respondent has a relative preference for the regional variety in the informal context and shifts towards a preference for SBD in the formal context; (2) the

respondent has an overall preference for the regional accent in both contexts, but that preference diminishes in the formal context; (3) the respondent has an overall preference for SBD, but that preference is stronger in the formal context. All three scenarios will be considered together and regarded as the hypothesized context effect. The reason why we are not distinguishing between the three scenarios in the analysis is that they are defined in reference to a D score of zero. This is, however, not an absolute bench-mark or neutral reference point, hence it should be approached with some caution.

3 Results

3.1 Implicit attitude measurement

The mean E_{contex} scores per condition are visually summarised in Figure 7. In order to establish whether a context effect was measured and whether there were any mediating factors in the design, a linear regression model was built with the difference score between the formal and informal P-IAT ($E_{context}$) as the response variable and context type (background vs. slides), context order (formal-informal vs. informal-formal), block order (BO1 vs. BO2) and the interaction between the latter two as fixed effects (see Table 11)[29]. As sum coding was used in the model, the intercept reflects the grand mean. The fact that the intercept is not significant indicates that no overall context effect could be observed. Furthermore, the type of context presentation did not have a significant effect either.

[29] The model was checked for influential cases and outliers. No data needed to be removed from the analysis. Additionally, the assumption of no multicollinearity was met as well as those regarding the residuals of the model.

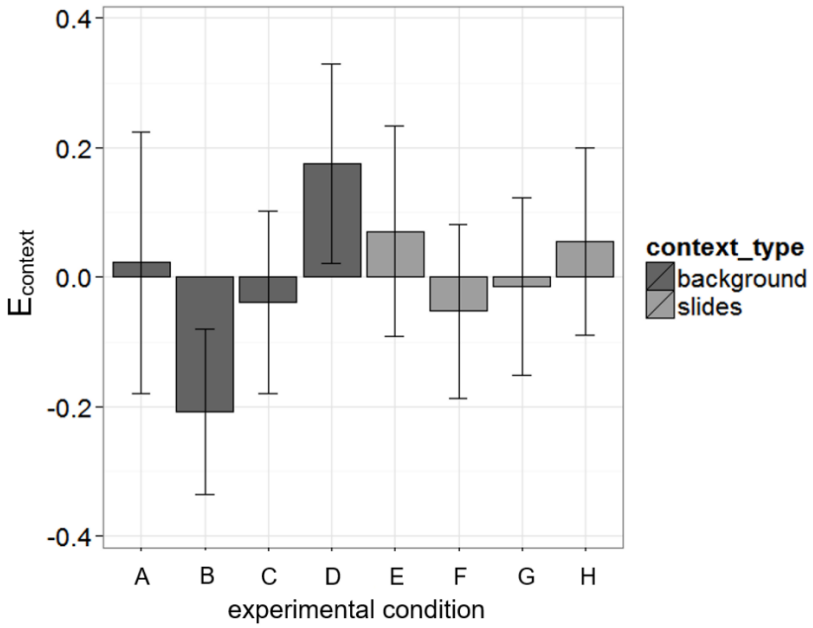


Figure 7 - Mean difference scores ($E_{context} = D_{for} - D_{inf}$) plotted per experimental condition (see Table 9 for an overview of the conditions). Positive scores indicate behaviour in line with the hypotheses. Negative scores indicate the opposite trend.

The regression model shows a significant interaction between context order and block order. Mean positive difference scores, which indicate participants behave as predicted by our hypothesis (i.e. that formal contexts increase preference for SBD and informal contexts increase preference for the regional variety), are only observed when block order 1 (BO1) is combined with formal-informal context order or when block order 2 (BO2) occurs with informal-formal context order (see Figure 8 for a visualisation of the interaction effect. Alternatively, the interaction is also visible in Figure 7 and summarised in Table 12). This pattern of block orders and context orders corresponds to the cases where the combination of block order and the context offered in the first P-IAT aligns with our hypothesis. As explained above, the IAT is known to produce lager effects if the first block is the congruent block, i.e. the block that is in agreement with a participant's attitudes. In this experimental design, block order 1 (BO1) is the congruent block order for participants who prefer SBD over the regional variety. According to our hypothesis, this situation particularly matches formal contexts. Conversely,

block order 2 (BO2) starts with the congruent block if the Limburg accent is preferred over the standard one, which we predicted to be particularly the case in informal contexts. In other words, we do observe the predicted context effect, but only if all factors in the experimental design create the most favourable conditions to do so.

predictor	estimate	*p*	
intercept (grand mean)	0.001	.97	n.s.
context order			
formal-informal	-0.042	.10	n.s.
context type			
background	-0.013	.61	n.s.
block order			
BO1	0.009	.74	n.s.
context order x block order			
formal first x BO1	0.079	< .01	**

model statistics: Adjusted R^2 = .056; F = 3.22, p = .015

significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 'n.s.' 1

Table 11 - Summary of the linear regression model with difference scores between the formal and informal P-IAT ($E_{context}$) as the response variable. Positive difference scores indicate behaviour in line with the hypotheses. Negative scores indicate the opposite pattern.

	BO_1	BO_2
formal - informal context order	positive $E_{context}$ = hypothesis	negative $E_{context}$ ≠ hypothesis
informal - formal context order	negative $E_{context}$ ≠ hypothesis	positive $E_{context}$ = hypothesis

BO1: first experimental block has SBD/positive & CBD/negative key mapping
BO2: first experimental block has CBD/positive & SBD/negative key mapping

Table 12 - Summary of the interaction between context order and block order in the regression model presented in Table 11

In order to make sure a potential context effect was not masked by fatigue or practice effects due to the participants having to complete two subsequent P-IATs, we analysed the data of the first P-IAT separately. This means the experimental design was reinterpreted from a within subject context manipulation to a between subject one. No evidence was found to suggest that that was the case.

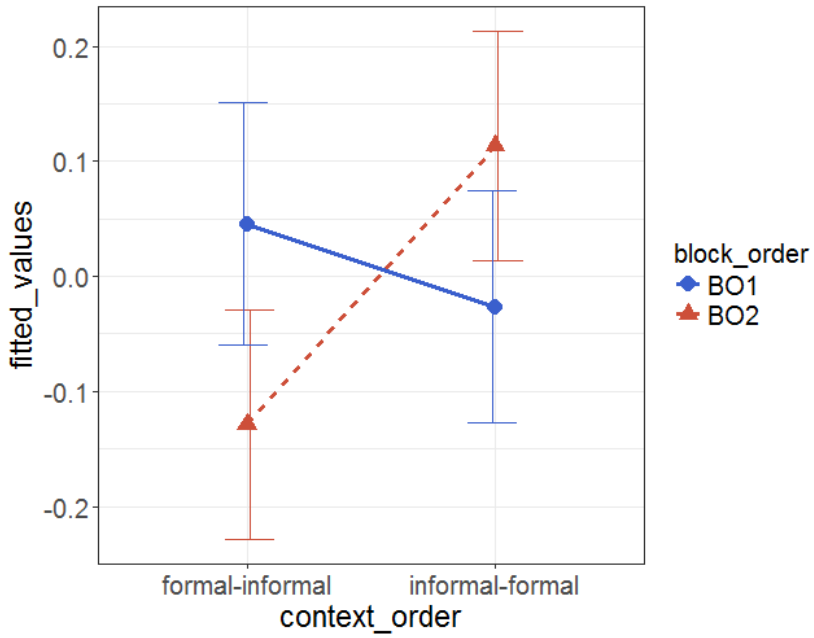


Figure 8 - Effects plot showing the interaction of context order and block order (Chapter 3)

3.2 Explicit attitude measurement

Prior to the analysis of the explicit attitude measurement, we discarded the data obtained from participants who we suspected may have misinterpreted the rating task. These participants showed discrepancies in their ratings on the relative and absolute scales which seemed to indicate that they confused the 'negative'-'positive' labels on either end of the absolute scales. This suspicion was further backed by the fact that a few participants apologized for handing in a rather messy form after completing the experiment, as they had corrected their answers after misreading the scales. Of course this is a post hoc interpretation and we cannot be certain it is correct. Hence, we concluded it

would be safer not to include the data in question in the analysis. This lead to the exclusion of the data provided by 23 participants. Additionally, one participant forgot to complete the explicit rating task, leaving a total of 128 participants for this analysis.

As explained above, both the SBD and Limburg accent were rated for formal and informal contexts in two ways: (1) comparing both varieties on a relative 7 point rating scale and (2) evaluating each variety separately on an absolute 7 point rating scale (see Appendix A). Contrary to the implicit measurement, the explicit rating task shows a clear context effect, both for the relative and absolute rating scales (see Figure 9). When asked to rate both accents compared to each other, participants strongly preferred the standard accent in formal contexts (M = 6.27, SE = 0.08) and show a slight preference for the regional accent in informal contexts (M = 3.14, SE = 0.14, cf. the bottom row of boxplots in Figure 9). A paired t-test with Bonferroni correction indicates that this context effect is significant and substantial ($t(127)$ = -22.34, $p < 0.001$, r = .89).

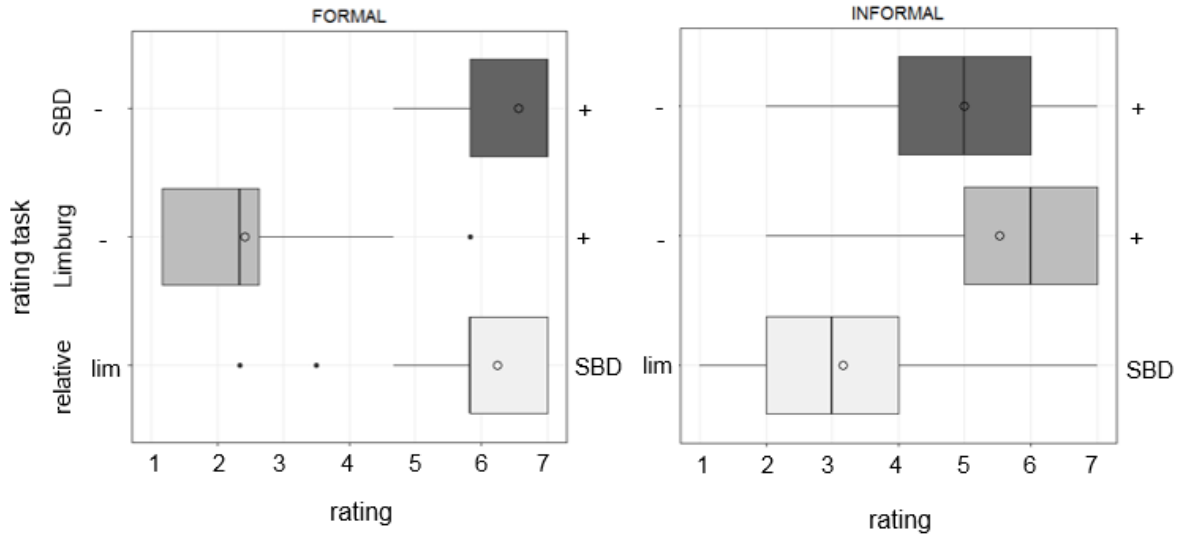


Figure 9 - Boxplots summarising the results from the absolute (two upper boxplots) and relative (bottom boxplot) rating tasks in formal (left plot) and informal (right plot) contexts. Lower scores on the relative scales indicate a preference for the regional accent, higher scores for the standard accent. For the two absolute scales, lower scores correspond to more negative evalutations of the respective varieties, higher scores to more positive ones.

Turning to the absolute ratings of each variety separately, we can observe a different pattern in the formal and informal context (see the upper two rows of boxplots in Figure 9). In a formal context, SBD is viewed very positively (M = 6.54, SE = 0.05), while the regional variety is evaluated extremely negatively (M = 2.44, SE = 0.11). Again this is a large and significant effect ($t(127)$ = -34.13, $p < 0.001$, r = .95) that is in line with what we predicted to find. In the informal context, the regional accent receives the highest mean score ($M_{informal}$ = 5.51, SE = 0.12 vs. M_{formal} = 4.98, SE = 0.12, $t(127)$ = -3.11, $p < .01$, r = .27). However, the standard accent is rated quite positively as well, which does not align with what we hypothesized.

3.3 Correlation analysis implicit-explicit attitude measurement

To study the correlation between the implicit and explicit attitude measurement, Spearman's rho was computed and tested per experimental condition for each explicit rating scale with the D score of the corresponding P-IAT. For instance, for condition A, we first calculated the correlation coefficients between each of the three rating scales (two absolute and one relative) in the formal context and the D score of the first P-IAT, which corresponded to the one presenting the formal context in this condition. Next, we calculated the correlation between each of the three informal explicit rating scales and the D scores for the second (informal) P-IAT.

Overall, the correlations between the two types of measurements were weak to moderate and the majority were not statistically significant. Hence, it is hard to draw any conclusions or make generalizations about the relationship between the implicit and explicit attitude measurements. However, there was a trend for the relative explicit ratings and the absolute explicit ratings of SBD to be positively correlated with the D scores, while the explicit rating of the regional accent tended to be negatively correlated with the implicit measure. In other words, the implicit and explicit evaluations point in the same direction: a stronger explicit preference for the standard accent was associated with a stronger implicit preference for this variety and similarly for the regional variety. This was the case regardless of context. Note though that these are mere trends as in most cases the correlations were not strong or significant.

4 Discussion

Why did we find a context effect in the explicit attitude measurement, but only one in interaction with context order and block order in the implicit attitude measurement? In this section we will discuss two potential avenues of explaining the results reported above. First, possible methodological issues will be raised, followed by a suggestion to review the hypothesis we started out with.

The absence of a context effect in the P-IAT experiments may be the result of methodological choices in the experimental design. One locus for such issues are the experimental stimuli. Although we did some preliminary testing on the selection of stimuli representing the formal and informal context, it is possible these were not suitable for evoking the desired contexts. For instance, we deliberately avoided pictures of Belgian news studios for the formal context in order to avoid biasing the results due to confounding associations with certain newsreaders or TV channels. Yet, the alternative of using foreign news studios may not have led to the evocation of the formal context we hypothesized the participants would associate with SBD. On the side of the informal stimuli, we aimed to evoke situations where participants interact with local friends and family during drinks or dinner. However, some participants may have activated experiences of informal contexts with friends from other regions or family that lives in a different area of the country. In those cases, the use of a regional accent that does not match the accent of the interlocutors may not necessarily be viewed positively, despite the informality of the situation. Other possibilities are that the way of presenting the stimuli was not optimal and as a result the desired contexts were not activated. Maybe context is better evoked using stimuli of a different modality (e.g. the theme tune of the TV news). Or perhaps a more successful approach is to present the context through a separate task preceding the experiment rather than by showing pictures during the P-IAT. That way participants may be forced to actively engage with the situational contexts under study, which would rule out the possibility that some participants ignored the context pictures as they were not relevant for the

task[30]. Having participants engage more actively with the contextual stimuli may also lead them to evoke personal experiences which could help to activate the desired situational contexts. An approach like this would be more similar to the method employed in Wittenbrink et al. (2001) where respondents were asked to write a text about a video evoking a certain setting before proceeding to the IAT.

A second way to explain the absence of a context effect is that our initial hypothesis was too coarse to capture the more fine-grained reality of the social meaning of the language varieties under study. We suspect there may be an asymmetry in the context effect. It could be the case that participants are ambivalent to both varieties in both contexts. The regional variety allows the speaker to better express their identity, yet it is not very 'proper', while the standard does offer the potential to sound more competent/prestigious, but may come across as boring or lacking any individuality. Yet, this ambivalence may be stronger in the informal context than in the formal context where the standard is almost exclusively viewed positively. This situation can clearly be observed in the outcome of the explicit measurement where SBD was perceived positively as well as CBD in the informal context.

In addition, the context manipulation may have a different impact on participants with a different degree of 'norm sensitivity'. It could be the case that people who care more about what is expected by a prescriptivist perspective on language variation where each variety is strictly confined to its own domain, are more susceptible to the context cues in the experiments. This mindset can also be interpreted as a form of ambivalence: these participants hold both positive and negative attitudes towards a variety (depending on whether it is used in the appropriate context). Psychological research on ambivalence has found that ambivalent people tend to draw more on contextual cues, as these may help them to resolve their ambivalence in a specific situation (Petty & Briñol 2009). This supports our hypothesis about the mediating role of norm sensitivity. Unfortunately, our study did not include a measure of norm

[30] Note though that in the condition with the slides, participants were explicitly instructed to pay attention to the pictures.

sensitivity. Our best option to get a first idea of whether norm sensitivity/ ambivalence does indeed mediate the context effect in our P-IATs was to take the explicit ratings as a proxy. Participants were divided in three groups based on how much their absolute explicit ratings diverged for each variety between the two contexts. The first group consisted of non-ambivalent participants (difference score between the absolute ratings in each context of 0 or 1). The second set of participants were highly ambivalent (ratings that are 4 or more points apart between the formal and informal setting). The third group of participants lies in between these two: they are ambivalent to some extent, but not as outspoken and the second group. Using these criteria, each participant received a qualification as not/moderately/highly ambivalent for the regional variety and not/moderately/highly ambivalent for SBD[31].

Ambivalence to the regional variety does not seem to have an impact on the context effect in the P-IATs. Ambivalence towards the standard variety, however, mediates the context effect as expected: participants who are highly ambivalent towards SBD exhibit the predicted context effect, while the other groups do not (see Figure 10). Adding ambivalence towards SBD and Limburg accented speech to the regression model for the implicit measurement shows a significant main effect for ambivalence towards SBD, but not towards the regional variety. Ambivalence towards SBD contributes significantly to the model ($F = 7.79$, $p < .001$) and its addition improves the amount of variance explained by the model from adjusted $R^2 = .04$ to adjusted $R^2 = .15$. Note though that the group of participants that are classified as highly ambivalent is rather small (N = 10). Hence some caution is warranted regarding the

[31] Admittedly, the explicit measurement is not a measure of norm sensitivity, and this grouping of participants rather indicates their context sensitivity. Yet, we are assuming here that participants who are the most norm sensitive/ ambivalent, will be the most sensitive to differences in situational context. Hence, we take the latter as a proxy for the former. There is a degree of circularity in this approach, but within this study, it is the best approximation we have at our disposal. In that respect we emphasize that we do not consider this as more than a mere exploration. Future research will have to include more valid ways of measuring norm sensitivity independently of context sensitivity to find out whether norm sensitive participants' implicit language attitudes are mediated through context more strongly than those who are not or less norm sensitive.

observations reported above and more research on the relationship between language attitudes and ambivalence is needed before drawing any conclusion.

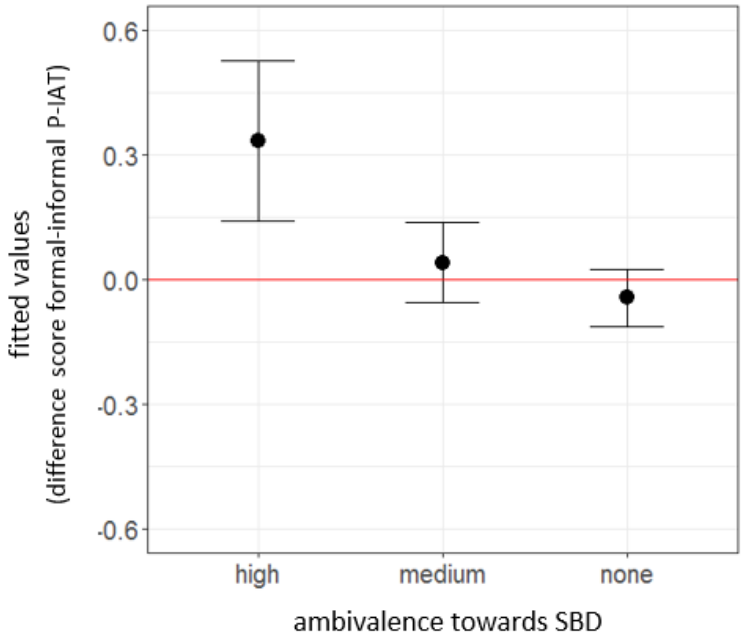


Figure 10 - Effects plot showing the context effect ($E_{context}$, i.e. the difference score between the formal and informal P-IAT) for participants with different degrees of ambivalence towards SBD

Given that our experimental design was based on an initial hypothesis that may have been too crude, it is possible that we were not able to measure any asymmetric context effects. It may be necessary to refine the experimental design, so we can test what we think could be a more accurate version of our hypothesis. One way of improving the design could be to include a baseline context in addition to a formal and informal context. This baseline would have to be neutral regarding situational formality. Measuring attitudes in this condition, would then provide a benchmark against which the impact of adding a formal or informal context to the experiment can be described more precisely. Note though, that operationalizing a neutral context may not be straightforward and that the addition of a third context type requires a more complex experimental design. That is to say, completing three P-IATs in a row may well be too exhausting for participants. Additionally, several measures for

norm sensitivity should be introduced as well, since norm sensitivity may influence the extent to which participants take into account contextual cues as well as determine whether context is an important mediating factor in their attitude towards a regional vs. a standard accent. Something worth considering also is that the IAT paradigm may not provide the most suitable methods to study context effects on the social meaning of language variation. Although, of course, our experiment can in no way prove that the IAT is insensitive to context – and previous social psychological work has shown that it is sensitive to context, the study does give an indication that it is not straightforward to successfully study the influence of situational context on language attitudes within the IAT paradigm.

As a final consideration about why we could only find a context effect on implicit attitudes in interaction with context order and block order, it is of course possible that situational context has no influence on these attitudes, hence making it impossible to measure it. However, taking into account a number of studies linking the production of Dutch language variation to situational context (see Section 1.3), combined with our own results for the explicit measurement that show a clear context effect[32], this is not a likely explanation. Yet, perception research on this topic could certainly benefit from more insights into current production patterns: is there still a strict distinction between the use of regionally and standard accented speech in different informal and formal settings? Is there evidence for the often heard statement that non-standard usage is creeping into contexts that used to be the sole domain of the standard variety? This would allow research on the social meaning of language variation to be more confident/specific in building hypotheses and designing suitable experiment designs. Specifically from the perspective of exemplar theory, production research on the co-occurrence of language and certain contextual features could contribute to a better understanding of which associations we can expect between language and social meaning. Likewise, production research on this topic could benefit from

[32] Note, though, that a context effect on explicit attitudes does not guarantee there should be one on the implicit level.

more studies into the social meaning of language variation that take into account situational or other types of contexts in that these studies may provide explanatory potential for certain patterns of variation and change. Consequently, we would like to end this discussion section with a plea for more research on the role of contextual factors in both the production and perception of varieties of Dutch in Belgium.

5 Conclusion

The Personalized Implicit Association Test (P-IAT), a method relatively new to sociolinguistic research, presents linguistic stimuli void of context: participants only receive labels for the language varieties or variants under study and very brief auditory (or written) exemplars representing these varieties or variants. Given the crucial influence of context cues on the perception and evaluation of language variation (e.g. Campbell-Kibler & McCullough 2015), this study aimed to explore the potential of the P-IAT to incorporate contextual cues by introducing situational context into the experimental design. We hypothesized that Limburgish participants would prefer their own regional variety in informal settings, but standard accented speech in formal contexts. Regardless of whether the visual context cues were presented during the P-IAT trials or in between blocks of trials, a context effect was found only in interaction with context order and block order. An explicit attitude measurement, on the other hand, confirmed the hypothesis, but showed very positive attitudes towards both the regional *and* the standard accent in informal contexts. Multiple explanations for the absence of an overall context effect in the P-IATs can be put forward. Firstly, there may have been issues in the experimental setup, which prevented the actual activation of the situational contexts that we aimed to evoke. Secondly, our initial hypothesis may have been too crude. More production research on the link between situational context (and other types of context) and Dutch language variation will hopefully provide useful input to base more fine-grained hypotheses on in the future. Finally, ambivalence and norm sensitivity were discussed as potential sources of individual variation between participants' performance on the P-IATs. We strongly suggest that

both factors be taken into account in future research on the influence of context cues on (Belgian Dutch) language variation.

To conclude, we would like to encourage sociolinguists to try and include context features in their experiments on language attitudes, however challenging that may be, in order to obtain results that are more ecologically valid. In addition, more research on how social meaning of language variation interacts with any type of contextual information will further add to our understanding of sociolinguistic cognition.

Chapter 4

Measuring the social meaning of Standard and Colloquial Belgian Dutch with the Relational Responding Task

1 Introduction

This chapter reports a study on the social meaning of two varieties of Belgian Dutch: Standard Belgian Dutch (SBD) and Colloquial Belgian Dutch (CBD). In the previous two chapters, we measured attitudes towards SBD and three regional varieties of CBD in terms of valence. The study in this chapter focuses on more specific aspects of the social meanings of these varieties. It investigates whether the standard variety is associated with prestige evaluations, while the colloquial variety is perceived in terms of dynamism, as suggested by a number of recent studies (e.g. Grondelaers & Speelman 2013; Impe & Speelman 2007). In that sense, this chapter is more closely related to traditional attitude research that strongly focuses on attitude dimensions such as prestige.

Parallel to the previous two chapters, the study consists of a measurement of implicit attitudes and one of explicit attitudes. However,

unlike Chapters 2 and 3, a new implicit attitude measure was employed, namely the Relational Responding Task (RRT). We will start the chapter with a brief description of the linguistic background of the study (Section 1.1) and then move on to discuss how and why the RRT was used for this study (Section 1.2). In the light of implementing the RRT as a language attitude measure, special attention will be given to the conceptualisation of language varieties through the use of labels and exemplars (Section 1.3; see Chapter 1 for a general discussion of this issue relating to the IAT and AP, as well as the RRT). Section 1.4 will bring these three aspects of the study together in specific research questions and formulate corresponding hypotheses. In the rest of the chapter, we will outline the study's experimental design (Section 2) followed by a summary of the main findings (Section 3) and conclude with a discussion of the results in Section 4.

1.1 From diglossia to diaglossia in Flanders: The perceptual perspective

To frame the study reported in this chapter, we need to go back to what was explained about the linguistic situation in Flanders in Chapters 2 and 3 and supplement this picture with a diachronic perspective. In the latter chapter, we presented the current stratification of Dutch in Flanders taking a language production perspective. Now, we will discuss how this current situation came about and what the perceptual counterpart of this change may involve. To that end, we will concentrate on a few of the studies that were mentioned in Chapter 2 where the literature on language attitudes in Flanders was reviewed. Against this background, we can then indicate how this study aims to contribute to a better understanding of the current stratification of Dutch language variation in Belgium.

In Chapter 3, the current linguistic situation in Flanders was described as a diaglossic one (Grondelaers & Van Hout 2011; Geeraerts & Van de Velde 2013; Ghyselen 2016). This diaglossic situation is the result of a change in the past 60 years or so, which had a diglossic starting point (Geeraerts 2017). This change is summarised in Figure 11. The cone on the left represents the diglossic situation which comprises two stratificational levels: the standard variety and

the base dialects. The size of the ellipses in the cone symbolizes the amount of variation within a stratificational level, while the distance between the ellipses represents the linguistic distance between the varieties. Moving from the older diglossic situation to the current diaglossic one, we can see that the bottom of the cone has shifted upwards somewhat. This upward shift represents dialect loss, a phenomenon which characterizes most of Flanders (with the exception of the westernmost periphery, Ghyselen 2016). A second difference between the two cones, and the one most crucial to our discussion here, is the appearance of CBD in the second cone. In the current situation of diaglossia, we no longer have two separate levels in the stratification, but rather a continuum where CBD fills up the space between the base dialects and SBD. A number of factors have been cited that may have played a role in this change from diglossia to diaglossia and the rise of CBD. Some authors refer to dialect loss (Willemyns 2007), others to tendencies of democratization and informalisation in postmodern society (Grondelaers & Van Hout 2011; Geeraerts & Van de Velde 2013) or to increased self-confidence in Flemish society (Grondelaers & Van Hout 2011).

Now, the question we are interested in here is: what are the perceptual correlates of this change in production? In a review of older (i.e. 1970s and 1980s) language attitude research in Flanders, Grondelaers (2013) indicates that dialects are judged suitable only for informal and homely situations, and, in some studies, they elicit evaluations of social attractiveness. The standard variety, by contrast, is perceived as prestigious. What we want to study here is whether these associations have changed in the light of the arrival of CBD and what the social meaning of the latter variety is. As discussed in Chapter 2, there is some recent work on this question showing that SBD has retained its prestige evaluations and is still perceived as a language variety signalling status and competence (e.g. Impe & Speelman 2007; Van Gijsel et al. 2008; Grondelaers & Speelman 2013; Grondelaers & Van Hout 2016). CBD is proposed to carry indexical meanings of dynamism, which is defined as a sort of cool associated with the media, and characterizes speakers as trendy and assertive. As for the bottom of the cone in Figure 11, few recent studies focus on the current evaluations of the traditional base dialects and neither will we in the study

reported below. Yet, based on findings reported in Vandekerckhove & Cuvelier (2007), Impe & Speelman (2007) and Lybaert (2014), we hypothesize that the local dialects have retained their indexical meaning in terms of informality, friendliness and homeliness. Of the studies mentioned above, Impe & Speelman (2007) is the only one to present recent research that reports on all three layers of social meaning (prestige, dynamism and homeliness) within one study. Using a verbal guise technique, they showed that the most standard guises in their experiment received prestige evaluations, while the guise that was considered more dialectal was associated with familiarity and homeliness. The guise that included the most CBD features was linked with dynamic characteristics. In our study, we aim to gather more evidence on the social meanings of SBD and CBD, and by doing so we hope to contribute to an increased insight into the current language dynamics in Flanders.

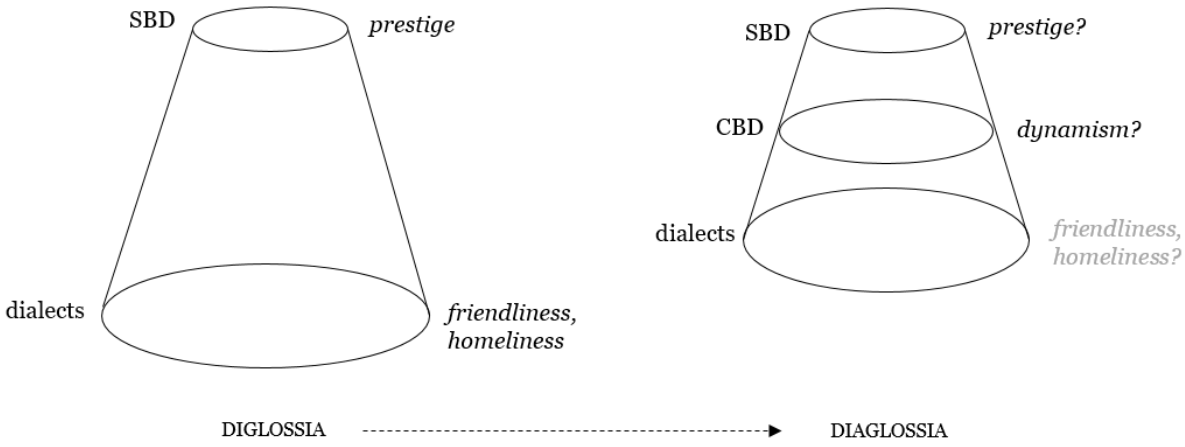


Figure 11 - Schematic representation of the change from diglossia to diaglossia in Flanders, based on Geeraerts (2017). The (proposed) social meanings associated with the various language varieties are printed in italics on the right of each cone.

1.2 How to measure the social meanings of SBD and CBD?

The dynamism associations with CBD reported in previous research have been defined as 'covert', 'subconscious' or 'deep' evaluations and, because of that, are deemed to be particularly difficult to measure using direct attitude measures, or even using traditional indirect methods (Grondelaers & Speelman 2013;

Grondelaers et al. 2016). Grondelaers & Speelman (2013), for instance, relied on the speaker evaluation paradigm to measure the social meaning of CBD. In the study reported below, we attempt to provide additional evidence for the dynamic indexicality of CBD by using a social psychological implicit attitude measure: the Relational Responding Task (RRT).

The RRT is a recently developed implicit attitude measure that allows researchers to study beliefs at the implicit level (De Houwer et al. 2015). Beliefs[33] are propositions that contain information on how concepts are related to one another. As explained in Chapter 1, social psychologists define beliefs as constructs that rely on propositional rather than associative processes. Traditionally, implicit attitudes have been linked to associative processes, while explicit attitudes and beliefs were assumed to rely on propositional reasoning (Hughes et al. 2011; De Houwer 2014). This idea of two types of processes underlying two types of attitudes lies at the core of dual-process models of attitudes that have been dominant in social psychological attitude research in the past decade or so (e.g. Gawronski & Bodenhausen 2006). Yet, recently, there has been a paradigmatic shift in cognitive models of evaluation leading to single-process models that view associative processes as a product of propositional ones (Hughes et al. 2011: 478). This turn to single-process models of evaluation fits in with a similar shift in conceptualising human learning and cognition at large (Hughes et al. 2011). This new perspective on the cognitive processes underlying evaluation suggests that not only associations, but also propositions can be activated automatically (i.e. fast and unintentionally, De Houwer 2014). In response to this theoretical shift, measures have been

[33] As discussed in Chapter 1, we will not adhere to a sharp distinction between attitudes and beliefs. Firstly, this distinction is often hard to maintain (e.g. Albarracín et al. 2005). Secondly, by not insisting on a sharp divide between beliefs and attitudes, we follow a recent tendency in linguistic attitude research to define language attitudes broadly (cf. Preston's (2011) 'language regard'). And thirdly, if, as we argue in this section, beliefs are assumed to be able to operate at the implicit level, just like attitudes, the remaining difference between the two constructs lies in the degree to which the link between the concepts relevant to the evaluation is specified. When this difference is relevant, it will be mentioned explicitly.

developed to study propositions at the implicit (i.e. automatic) level, one of which is the RRT.

In this study, we propose to use the RRT as an alternative to the speaker evaluation paradigm which has been the dominant indirect approach to the study of language attitudes towards SBD and CBD. The RRT shows some similarities to this traditional approach, yet differs from it in ways that may allow us to measure attitudes under automatic conditions. In the study reported below, we constructed our RRT experiment so that it would be structurally as similar to matched/verbal guise experiments as possible. Firstly, in our study participants listened to two audio samples representing SBD and CBD, just like in the MGT/VGT. These were recorded by the same speaker, but presented to participants as coming from two different speakers. Subsequently, participants completed two RRTs (one for prestige and one for dynamism) that measured the social meanings of the two varieties under study. The stimuli in the RRTs were statements that linked either prestige or dynamism traits to the language varieties (e.g. '[SBD] sounds more intelligent than [CBD]'). These traits were presented by adjectives, as is usually the case in MGT/VGT experiments. Note that the use of propositions as stimuli is one of the most important differences between the RRT and the IAT. The latter only measures the association between concepts without specifying how they are related (see Chapter 1).

The statements in our RRTs were formulated in a relative manner (e.g. '[SBD] sounds more intelligent than [CBD]' rather than '[SBD] sounds intelligent'). This relative phrasing, combined with the inclusion of two varieties in the RRTs, rather than one, has a double motivation. Firstly, speaker evaluation experiments usually include multiple languages or varieties, as was the case in previous studies on the social meaning of SBD and CBD (Impe & Speelman 2007; Cuvelier & Vandekerckhove 2007;Grondelaers & Speelman 2013). If we want to be able to compare our findings to those studies, our study should be structurally and conceptually similar (Payne et al. 2008; Hofmann et al. 2005a: 1380). So if previous work relied on within subject experimental designs where participants judge multiple varieties, our study should take a similar relative approach. Secondly, as explained in Chapters 1 and 2, we

believe that the evaluation of language is an inherently relative process. Hence, making explicit which concepts are compared in the experiment ensures a better controlled study.

Despite structuring our RRT experiment as much as possible as a traditional speaker evaluation experiment, there are of course many differences between the two approaches. One way in which they diverge is the precise measuring procedure. Whereas a MGT/VGT experiment usually features ratings on semantic differential scales, the RRT requires participants to categorise statements according to specific instructions. Respondents indicate whether statements are true or false. Yet, they do not react to the statements according to their own opinions, but they follow a rule that is specified in the instructions of the RRT. For instance, participants react to the statements *as if* they believe one language variety sounds more prestigious/dynamic than the other, rather than judging the statements based on what they actually believe about the varieties. What is measured is whether participants carry out the classification tasks correctly and how fast they perform them depending on the rule specified in the instructions. This procedure is designed with a view to measure participants' reactions under automatic conditions. This brings us to a second point on which the RRT and the speaker evaluation paradigm differ. Linguists have used the latter to measure what they refer to as 'covert' attitudes. As discussed in Chapter 1, covertness seems to be mainly defined in terms of unawareness: respondents are kept unaware of the fact that the researcher is trying to measure their language attitudes. The RRT, by contrast, aims to measure attitudes under conditions of automaticity, which means that it tries to measure fast and undeliberate evaluations. This may mean that participants are unaware of what is being measured, but it is not necessarily the case. Perhaps working with methods that aim for implicitness in the social psychological sense rather than covertness in the sociolinguistic interpretation will allow us to measure the social meaning of CBD more easily.

In addition to taking a new approach to the indirect measurement of the social meaning of CBD (and SBD) by using the RRT, we also included a direct attitude measure in our study in the form of a direct rating task. The reason for including a direct measure is that we wanted to verify whether it is indeed still

the case today that it is impossible to capture explicit dynamism evaluations of CBD using such a direct measure, as suggested by Grondelaers & Speelman (2013). Previous production research has shown that CBD is mainly associated with a younger audience (Van Gijsel et al. 2008) and used more often by younger speakers (Plevoets 2008). Some perception studies have also signalled an age effect. Vandekerckhove & Cuvelier (2007) found age differences in the evaluation of CBD in some of their experimental conditions whereby younger respondents were more positive towards CBD. Ghyselen (2009) reports an age effect too, but in the opposite direction with younger participants being less acceptant of CBD features. Note, though, that Ghyselen (2009) focused on a specific regional sample of participants and one particular morphosyntactic CBD feature which makes it harder to generalize her results. Given these findings, it may not be entirely implausible to measure explicit positive dynamism evaluations for CBD in younger participants.

1.3 What's in a name: The labelling problem revisited

In the first Chapter's comparative review of implicit attitude measures, it was pointed out that including auditory stimuli in a RRT is not as straightforward as in, for example, an IAT. Given that the RRT uses written statements as its stimuli, a different approach was necessary to include actual language use in the experiment. As a solution, two longer sound samples (ca. 30 seconds), each representing one of the varieties under study, were played before participants started the RRTs. The varieties presented in those audio samples could then be referred back to in the RRT's statements. In order to do so, the varieties needed to be labelled in one way or another, of course.

In Chapter 1, the issue of labelling language varieties in experimental attitude research was introduced. Social psychological research has demonstrated in the context of the IAT that the way a category is construed determines the attitudes that are measured (Lane et al. 2007). Both the selection of specific exemplars belonging to the category, as well as the choice of the category label are crucial for its construal. Linguistic studies have come to similar conclusions regarding language attitudes (Preston & Niedzielski 2013). For instance, Bishop et al. (2005a) and Coupland & Bishop (2007)

conducted a large-scale study on language attitudes in the UK. They used a 'conceptual method' in which they only presented their respondents with labels for different regional varieties. Their experimental design also included a number of alternative labels for similar varieties, such as 'Queen's English' or 'Standard English accent' for the standard variety, or labels implying a hierarchical categorisation, such as 'Welsh accent' vs. 'Cardiff accent' vs. 'Swansea accent' or 'London accent' vs. 'Cockney accent'[34] of which the former label is more vague and inclusive regarding for instance social class. Their results show that using different labels activated different social meanings of the regional varieties in their study, which lead them to conclude that '[c]onceptually based attitudes research can make visible how social attitudes and ideologies are category-bound' (Bishop et al. 2005a: 151). Another example of how the use of exemplars vs. categories can impact the results of a language attitude study can be found in Prikhodkine (2015). This paper focuses on attitudes toward varieties of French in Switzerland. The study shows that when respondents are presented with global category labels, only certain negative associations can be measured, while prestige evaluations linked to specific Swiss French features stay out of the picture. A final example, this time taken from an American context, is reported by Carmichael (2016). In her study about the influence of sociolinguistic knowledge on evaluating linguistic stimuli, Carmichael used both matching and mismatching labels for the linguistic stimuli in her matched guise experiment. She found that expectations linked to the label influenced how participants reacted to a speech sample.

In our study, we attempted to turn the labelling issue entailed by the RRT's procedure into an opportunity to gain more insight in how categorisation impacts on language attitudes. We did so by creating five between subject experimental conditions in which we referred to the SBD and CBD sound samples using a different set of labels (see Table 13). The labels were chosen so they would form a continuum from focus on speech to focus on speaker. In the two 'focus on speech' conditions, labels were used that named the language

---

[34] 'Cockney accent' was not included in the survey conducted by Bishop and colleagues (2005a), but it was in the Giles (1970) study which was partially replicated in the Bishop et al. (2005a) study.

varieties under study. In both of these conditions, SBD was referred to as *AN* which is an abbreviation for *Algemeen Nederlands* ('General Dutch') and a term often used by non-linguists to refer to the standard variety. In the first 'focus on speech' condition, CBD was labelled as *tussentaal*, a term for CBD used by both linguists and laymen (Lybaert 2014). The second focus on speech condition referred to CBD as dialect, which is a misnomer as the audio sample did not contain dialectal speech. Although our study will not allow us to investigate the exact cognitive processes involved in this mismatch between exemplars and category label, we deemed it worthwhile to explore the impact it would have on the social meanings associated with CBD in this condition.

overtness continuum	condition	label SBD	label CBD	# participants
FOCUS ON SPEAKER	TV hosts	*Bart Schols*	*Otto-Jan Ham*	65
	first names	*Pieter*	*Jonas*	29
		Jonas	*Pieter*	23
				52
FOCUS ON AUDIO SAMPLES	reference to audio samples	*fragment 1* ('sample 1')	*fragment 2* ('sample 2')	160
FOCUS ON SPEECH	language labels (accurate)	*AN* ('general Dutch')	*tussentaal* ('in between language')	63
	language labels (partially inaccurate)	*AN* ('general Dutch')	*dialect* ('dialect')	51

Table 13 - Experimental design with participant numbers (Chapter 4)

The two 'focus on speaker' conditions did not use language labels to refer to SBD and CBD, but named the supposed speakers of the samples. In the first 'focus on speaker' condition, the audio samples were referred to using *Pieter* and *Jonas*, two first names that were among the 20 most common given names for boys in the generation of our test sample[35] (see Section 2.2). We speculated that (social) connotations of first names, as well as knowing someone with a certain name could potentially impact the results of our study. Hence we decided to use two names that were as connotatively neutral/vague as possible (based on our own intuitions), although we realize that no first name can ever be devoid of connotations. We also opted for highly frequent names, so the majority of our participants would know multiple people with those names[36].

For the second pair of 'focus on speaker' labels, two names of well-known TV hosts were used: *Bart Schols* for the SBD guise, *Otto-Jan Ham* for the CBD guise. Both men are talk show hosts of a similar age. Their respective shows discuss current affairs and are broadcasted on the same TV channel (Canvas) run by the public broadcasting company. Crucially, the two shows differ with respect to how serious their take on the news is. Bart Schols presents *De Afspraak* which is a serious and more traditional talk show, while Otto-Jam Ham takes a frivolous approach to what has been happening in the news in his show *De Ideale Wereld*. Moreover, Bart Schols has always worked for the news service of the public broadcasting company, while Otto-Jan Ham has a past as a presenter at a trendy radio station (Studio Brussel) that mainly targets a young and fashionable audience. Given this background, these men's profiles

---

[35] Based on statistics published by the Belgian Government and available on http://statbel. fgov.be/nl/modules/publications/statistiques/bevolking/bevolking_-_voornamen_van_de _ pasgeborenen_1995-2014.jsp.

[36] We decided to control for the factor of knowing someone by these first names by choosing highly frequent names. The other option would have been to use made up names or names that are very infrequent, so hardly any respondent would know anyone by that name. Because we expected participants may have reacted more strongly against names that are unfamiliar to them, we deemed working with highly frequent names the better of two non-ideal options. Investigating attitudes towards a speaker depending on whether the listener knows someone with the same name as the speaker or does not know the name at all could be a study in its own right.

were deemed to match our distinction between prestige (Bart Schols) and dynamism (Otto-Jan Ham) rather well.

The fifth and final condition in the study does not refer to either the speech or the speaker of the audio samples, but to the samples themselves. It included the labels *fragment 1* ('sample 1') and *fragment 2* ('sample 2). We considered them to represent a middle way between the two 'focus on speech' and the two 'focus on speaker' conditions.

These five conditions form a continuum in terms of 'overtness' in the linguistic sense of the word (i.e. awareness). The 'focus on speech' conditions make the participants highly aware of what the study is about, while the 'focus on speaker' conditions draw less attention to language. In that sense, the 'focus on speaker' conditions bring the RRT closer to the traditional speaker evaluation experiments used in linguistics where keeping listeners in the dark about the aim of the experiment is crucial. Note, though, that even in the 'focus on speaker' conditions we did not actively try to mislead the respondents and divert attention away from language. In that respect, even our RRTs in the focus on speaker conditions do perhaps not qualify as measures of covert language attitudes in the linguistic interpretation of unawareness, as respondents may have been well aware of the fact that the experiment had something to do with language. However, as explained above, by using the RRT, we aim to investigate the social meanings of SBD and CBD under conditions of automaticity, which goes beyond awareness alone.

From a theoretical point of view, our manipulation of conditions with a view to gain more insight into categorisation mechanisms involved in the activation of social meanings of language variation fits in the scope of Cognitive (Socio)linguistic research (Kristiansen & Dirven 2008; Geeraerts et al. 2010; see Chapter 5). In Cognitive Linguistics, categorisation research occupies centre stage. The field sees linguistic meaning as a cognitive phenomenon (Langacker 1987). Individuals categorise experience as mental concepts which in turn are connected to linguistic forms. In that respect, categorisation processes are a crucial mechanism in language, just like they are in other fields of human cognition (cf. Lakoff's (1990) Cognitive Commitment; Dabrowska & Divjak 2015). Given the traditional dominance of semantic studies in Cognitive

Linguistics, the field has been primarily concerned with studying categorisation processes in referential meanings of language. However, we suggest that the social meaning of language can be studied in a similar way. As touched upon in Chapter 3, Cognitive Linguistics relies on prototypicality to make sense of the links between exemplars and concepts. We will return to this point in the next chapter.

1.4 Research questions and hypotheses

To summarise, this study aims to measure the social meanings of SBD and CBD in Flanders. We introduce the RRT as a novel tool to investigate implicit evaluations of those language varieties and complement this approach with a direct rating task. Based on the studies discussed in Section 1.1, we expect the following outcome:

- SBD is associated with prestige evaluations. This should be easily measurable on both the implicit (RRT) and explicit (direct rating task) level.
- CBD is associated with dynamism evaluations. Previous work (Grondelaers & Speelman 2013) suggests this will be difficult to measure on the explicit level, especially if CBD is contrasted with SBD (as is the case in our study). Hence, we may be more likely to measure it with the RRT than in the direct rating task.

That last point can be explained from two perspectives. According to Grondelaers & Speelman (2013), the reason why it may be difficult to measure positive attitudes towards CBD using direct methods is that the variety is still highly stigmatised in public debate and elicits strong negative reactions in the media. This stigmatization has to be understood against the background of the top-down standardisation history in Flanders which imposed an exoglossic Netherlandic standard. This repressive standardization history would explain why anything that does not correspond to SBD is harshly downgraded when respondents are explicitly asked about it, especially if directly compared to

SBD. The hypothesis that it may be more difficult to measure explicit dynamism evaluations towards CBD can also be sustained from a more cognitive perspective. Given that CBD is a newer variety than SBD (or dialect) in the stratification of language varieties in Flanders (see Section1.1), it may simply not be entrenched as a clearly delineated variety yet for many language users (cf. Plevoets 2008 or Ghyselen 2016 who emphasize the difficulty of delineating CBD, or Lybaert 2014 who reports that CBD is not perceived as a separate variety by all respondents in her study).

In addition to the hypotheses above, the study includes five conditions that refer to SBD and CBD using different labels. These labels are structured along an overtness continuum. Although our study will not allow us to draw any strong conclusions, the labelling manipulation was included as an initial and tentative exploration of how categorisation mechanisms interact with language attitudes.

Regarding the labelling manipulation, we expect that the 'focus on audio samples' condition will be the condition where it is the easiest for participants to make the connection between the sound samples they heard before starting the RRT, and the statements and rating scales in the RRT and direct rating task respectively. It is the most direct way of referring to the samples, as no additional categorisation or framing is suggested by the labels. Hence, we hypothesize to find clear prestige associations with SBD in both the implicit and explicit measurement, and at least in the former also dynamism associations with CBD.

In the 'focus on speech' conditions, there is the possibility that the labels activate additional social meanings other than the ones elicited by the sound samples themselves. This could especially be the case for the *AN-dialect* condition, as the label 'dialect' may evoke other associations than prestige or dynamism (cf. Section 1.1). Moreover, the two 'focus on speech' conditions represent the most 'overt' end of the continuum that structures our labelling choices. Given suggestions in the literature that dynamism associations with CBD are strictly covert, it may be more difficult to find evidence for these associations on the most overt end of our continuum even in the RRT. For the prestige dimension, on the other hand, we expect the presence of the explicit

AN label to lead to stronger prestige evaluations for SBD in these conditions than in the ones where the speakers are labelled rather than the speech.

We hypothesize to find clearer evidence of dynamism associations in the two 'focus on speaker' conditions, especially in the TV hosts condition as the two presenters were picked so their personality and show would support and reinforce the associations with the two dimensions under study. Note though, that just like with the 'focus on speech' labels, it is not unconceivable that either the first names or the TV hosts used in these conditions activate additional associations that are perhaps stronger and could compete with or even overrule the associations we are interested in, in this study. This may actually make it harder to measure the prestige and dynamism associations on both ends of the continuum.

Finally, we predict the labelling manipulation – which is a manipulation of degrees of overtness – to overall affect the dynamism ratings the most. The reason for this is that prestige evaluations of SBD have been measured in both overt and covert conditions in previous work, which has not been the case for CBD (e.g. Lybaert 2014 for overt prestige evaluations of SBD; Grondelaers & Speelman 2013 for covert prestige evaluations of SBD and covert dynamism evaluations of CBD). Hence, we expect the dynamism evaluations to be most sensitive to manipulations of degree of overtness.

2 Method

2.1 Overview experimental design

After listening to the sound samples containing SBD and CBD, participants started the study by completing the implicit attitude measurement. This measurement consisted of two RRTs: one measuring prestige evaluations, the other dynamism evaluations (see Section 2.3). The two RRTs were counterbalanced for order of appearance: half of the respondents started with the prestige RRT, the other with the dynamism RRT. After finishing the second RRT, respondents completed the explicit rating task (see Section 2.4). The experiment concluded with a brief questionnaire containing control questions as well as questions regarding basic demographic information (see Section 2.5).

Figure 12 summarises the flow of the study. All four stages of the study were completed online. The reason we chose to conduct the experiment through the internet was to recruit participants more efficiently. It allowed us to obtain a larger participant sample in a shorter amount of time.

Each participant was randomly assigned to one of the 5 conditions of the experiment (i.e. the different ways of labelling SBD and CBD, see Table 13). This was a between subject manipulation, so each participant only saw one and the same pair of labels during the study. As explained above, the labels were chosen so they formed a cline from focus on the speaker (TV hosts and first names) to focus on speech (*AN* vs. *tussentaal*/*dialect*), with focus on the audio sample itself as a middle category.

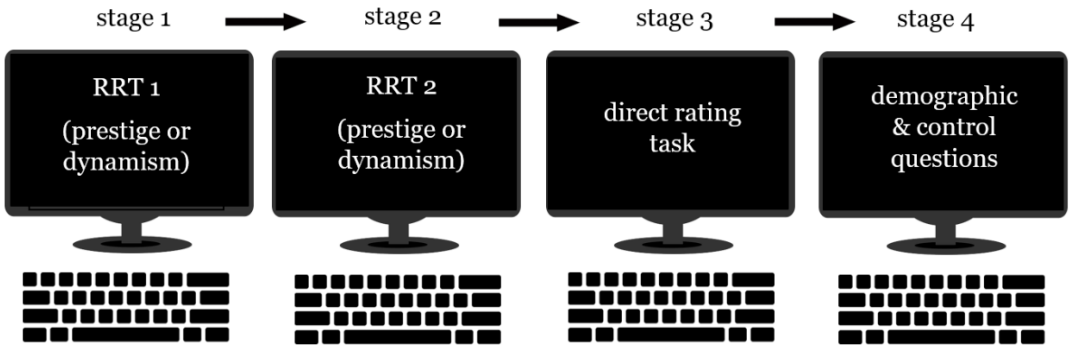


Figure 12 - Overview of the experiment (Chapter 4)

2.2 Participants

In total 430 participants completed the online experiment. Data from 39 participants were excluded from the analyses for one of three reasons. Firstly, participants who did not meet the demographic requirements for the study (i.e. they were not originally from Flanders or their native language was not Dutch) were removed from the dataset. Secondly, participants who reported considerable disruptions while they were taking the RRT were left out as well. The third criterion for exclusion was related to the response latencies and number of errors in the RRT. Participants with more than 10% fast responses (i.e. < 300 ms) and those whose number of errors was more than 2.5 standard

deviations higher than the mean number of errors were not included in the analyses reported below (De Houwer et al. 2015; Greenwald et al. 2003). Analyses were thus carried out on a sample of 391 participants (see Table 13).

2.3 RRT

2.3.1 Design and procedure

Before starting the first RRT, participants were asked to listen to two sound samples representing the two varieties under study, i.e. SBD and CBD. They were instructed that what was said was not important. Instead, they had to imagine what the two[37] people they heard were like and keep that in mind throughout the study. Once participants had listened to both audio samples, they were allowed to proceed with the two RRTs.

As explained in Chapter 1, one RRT consists of seven blocks in which participants categorise synonyms of true and false (the inducer stimuli) and classify statements as being true or false. The statements have to be judged based on a rule rather than on participants' own opinion. Respondents have to react to the statements *as if* they believe what the rule specifies (i.e. that one variety is more prestigious/dynamic than the other). Between the experimental blocks 3-4 and 6-7, the rule is reversed: if participants first had to react as if they believed SBD sounds more prestigious than CBD (blocks 3-4), they then had to respond as if they believed CBD sounds more prestigious than SBD in the next two experimental blocks (blocks 6-7), to use the prestige RRT as an example. This switch is first trained in a practice block (block 5) between the two sets of experimental blocks of the RRT. The precise structure of both the prestige RRT and dynamism RRT is summarised in Table 14 and Table 15 respectively.

Note that in both the prestige and the dynamism RRT, participants start with the congruent block. This is the block that – according to our hypotheses

[37] Note that participants were told they heard two different speakers, even though this was not the case. While preparing the experiment, we tested on a small sample of students whether this claim was convincing. We found no evidence to suggest that was not the case.

– represents the participants' belief. In other words, participants started with a rule stipulating that SBD sounds more prestigious than CBD in the prestige RRT and with a rule stating that CBD sounds more dynamic than SBD in the dynamism RRT. This block order was kept constant over all conditions. The reason why we opted for a fixed block order was that counterbalancing block order introduces additional error variance in the data, which makes it harder to establish correlations with other measures (De Houwer et al. 2015; Dewitte et al. 2017). Given that we were interested in the relationship between our RRTs and the explicit rating task, we deemed it more suitable to opt for a fixed block order in this study. By choosing the congruent block as the first block, we aimed to maximize the chance of measuring the hypothesized effect.

block	type of block	Task	# trials	stimuli
1	practice	categorising inducer trials	10	synonyms true/false
2	practice	categorising statements according to rule 1: [SBD] sounds more prestigious than [CBD]	20	statements about prestige
3	experimental	categorising inducer stimuli and statements according to rule 1	20	synonyms true/false + statements about prestige
4	experimental		20	
5	practice	categorising statements according to rule 2: [CBD] sounds more prestigious than [SBD]	40	statements about prestige
6	experimental	categorising inducer stimuli and statements according to rule 2	20	synonyms true/false + statements about prestige
7	experimental		20	

Table 14 - Structure of the prestige RRT (Chapter 4). [SBD]-[CBD] stands for one of the five pairs of labels used in the different conditions of the study.

block	type of block	Task	# trials	stimuli
1	practice	categorising inducer trials	10	synonyms true/false
2	practice	categorising statements according to rule 1: [CBD] sounds more dynamic than [SBD]	20	statements about dynamism
3	experimental	categorising inducer stimuli and statements according to rule 1	20	synonyms true/false + statements about dynamism
4	experimental		20	
5	practice	categorising statements according to rule 2: [SBD] sounds more dynamic than [CBD]	40	statements about dynamism
6	experimental	categorising inducer stimuli and statements according to rule 2	20	synonyms true/false + statements about dynamism
7	experimental		20	

Table 15 - Structure of the dynamism RRT (Chapter 4). [SBD]-[CBD] stands for one of the five pairs of labels used in the different conditions of the study

2.3.2 Materials

Three types of materials were used in the RRTs: audio samples, statements and inducer trials. The latter two were used as stimuli in the RRT, the former were presented to participants before each of the RRTs and in a shorter version between the blocks of the RRT as a reminder.

The audio samples were recorded by a professional radio presenter: a middle class man in his late twenties originating from a town north of Brussels in the province of Flemish-Brabant. We selected a professional voice working for the public broadcasting company in order to guarantee a pleasant voice

quality[38]. We made sure to select someone who is not well-known by a wide audience to avoid confounds in the experiment if the voice was recognised, and to make the cover story about each audio sample being a different speaker more plausible. In a series of pretests on the audio materials, no-one identified the speaker. The recordings were made in a radio studio to obtain optimal sound quality. Disfluencies and hesitations were edited professionally. The speaker was asked to produce standard language for the SBD guise. For the colloquial guise, he was instructed to speak as if he was interacting with friends from the same province. Both scripted and non-scripted recordings were made about a range of neutral topics. In the end, a pair of non-scripted samples about the same topic was selected as they sounded the most natural and authentic. The selected CBD sample contained a combination of colloquial features on all linguistic levels (i.e. phonetic, lexical and morphosyntactic) so the recording would come across as harmonious[39]. The full versions of the audio samples, played before each RRT, lasted approximately 27 seconds. Their short variants, played between the RRT blocks, were 9 seconds long.

In this study, a Brabantic variety of CBD was used. As mentioned in Chapter 2, some linguists believe that Brabantic features are spreading to other regional varieties of CBD (e.g. Goossens 2000). Considering this potential Brabantic expansion together with indications from previous studies that the Brabantic variety may be the most susceptible for dynamism associations (Impe & Speelman 2007)[40], we deemed that CBD with a Brabantic flavour would give us the best chance of measuring dynamism evaluations.

---

[38] TV or radio presenters working for the Flemish public broadcasting company (VRT) have to pass a three stage voice test, during which their voice quality is judged by a panel of experts (https://www.vrt.be/nl/jobs/acteren-presenteren/).

[39] Grondelaers & Van Hout (2016) introduce the concept of 'perceptual harmony'. They suggest that listeners are intuitively highly sensitive to language that has a combination of standard or non-standard language features that is out of balance given a certain context or interaction. Disharmonious language may engender strong (negative) reactions. We deemed that the non-scripted sample we selected contained a natural and balanced combination of CBD features. We did not find any evidence pointing to disharmony while pretesting the samples.

[40] Citing Willemyns (1979), Impe & Speelman (2007) explain their finding that the

The statements used in the RRTs were based on a number of traits that represent prestige and dynamism (see Appendix B). To come to a selection of traits, we started out with a set of adjectives used in previous research (Grondelaers & Speelman 2013), supplemented by adjectives we selected ourselves. In a next step, pretests were carried out on a small group of university students (N = 19) to validate these adjectives and verify whether any relevant aspects of prestige and dynamism were overlooked. During the pretests, respondents were asked to complete (1) a keyword task, (2) an adjective assignment task and (3) a speaker identification task. For the keyword task, respondents listened to the audio samples and jotted down as many words as they could to describe the speakers. The assignment task consisted of matching our selection of adjectives to the audio sample the participants thought was best described by that adjective. Finally, respondents were asked if they recognised the speakers. None of them correctly identified him. We then told the participants they heard Bart Schols and Otto-Jan Ham, the TV hosts whose names were used as labels for the audio samples in one of the experimental conditions. Mostly, participants' reactions were 'oh, yes, of course, now you mention it' or they indicated they already thought it must have been a TV host like Schols or Ham. The fact that the pretest group reacted positively and acceptant of the TV presenter guises increased our confidence that the guises were sufficiently convincing. The final selection of adjectives used for the prestige and dynamism statements was made based on keywords offered by participants and adjectives that were assigned unanimously in tasks (1) and (2) of the pretest. This final selection is summarised in Appendix B together with the inducer stimuli. The latter consisted of 10 synonyms of 'true' and 'false' and were taken from De Houwer et al. (2015).

---

central Brabantic CBD variety is the only variety linked to dynamism-like traits by suggesting that peripheral varieties of CBD, like CBD with a West-Flemish flavour, are more easily perceived by listeners as dialectal. As a result, these varieties of CBD are more easily associated with homeliness and intimacy, like the traditional base dialect usually are.

2.3.3 Analysis

For the RRTs, both response latencies and error information was collected. This data was used to calculate D scores employing the same D_6 algorithm as in the previous two chapters (Greenwald et al. 2003). The scores were calculated using the 'IAT' package for R (Martin 2015). We followed De Houwer and colleagues (2015) in only taking into account reaction times for the statements when calculating D scores.

2.4 Explicit rating task

2.4.1 Design and procedure

The explicit rating task consisted of 18 rating scales that were presented in a random order. For 18 prestige and dynamism traits, respondents were asked to indicate on a 7 point scale which speaker/variety/sample (depending on the condition) sounded the most like each trait (see Figure 13). They could do so by dragging a slider in the direction of the variety of their choice. The varieties were represented by the same labels that were employed in the RRT. Note that we used relative scales with each variety on one end of the rating scale. We opted for relative scales in order to maintain the relative nature of the RRT statements in the direct rating task, so the results would be as comparable as possible.

Before starting the rating task, participants received the explicit instruction that there were no rules to follow anymore and that this task was purely about their opinion. They were also asked not to think too long about their answers and to go with their first instinct.

2.4.2 Materials

The traits used in the rating scales were based on the traits used in the statements from the RRTs (see Appendix B). However, given the length of the study, we decided to use a slightly reduced selection of those traits in the explicit rating task in order not to discourage participants from completing the

entire study (e.g. 'more popular' and 'less popular' from the statements were only repeated in the rating task as 'popular').

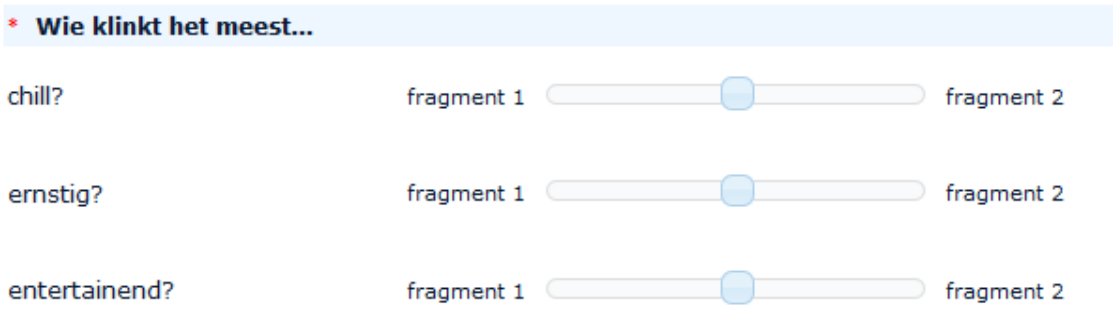


Figure 13 - Partial screenshot of the explicit rating task from Chapter 4 showing three rating scales for the traits 'cool', 'grave' and 'entertaining' (condition: 'focus on audio samples')

2.4.3 Analysis

To analyse the data from the rating task, we applied factor analysis. This is a statistical dimension reduction technique that allows one to verify whether the rating scales can be grouped according to underlying dimensions. This makes it possible to check whether our scales indeed group as a prestige versus a dynamism dimension and, if this is the case, whether SBD loads positively onto the prestige factor, while CBD scores better on the dynamism factor, as we predicted.

2.5 Control questions and demographic information

After completing the two RRTs and the rating task, participants filled out a short form containing questions regarding basic demographic information and a number of control questions. The demographic data collection concerned participant age, gender, region of origin, level of education and job/studies. This information was collected to allow us to get an idea of the demographics and homogeneity of our sample, rather than for the purpose of testing for differences in attitudes based on social groups. Basic knowledge about the

social make up of our sample may help to interpret the results of the study in a more correct and nuanced way.

The control questions were divided into two groups: condition-specific and study-wide questions. The latter group contained questions about the circumstances of the experiment (e.g. whether participants were disturbed or whether they experienced problems hearing[41]/seeing parts of the experiment), as well as three questions concerning norm sensitivity. For this norm sensitivity measure, respondents rated to what extent they agreed with each of three statements: (1) 'Different types of language (e.g. standard Dutch, dialect, *tussentaal*) are suitable for different types of situations'; (2) 'As long as everyone understands what you mean, the type of language you use is not important'; (3) 'I like hearing different accents'. The decision to include a norm sensitivity measure in the study was the result of the findings from Chapter 3 (see Section 4 in that chapter). In addition to the questions about the circumstances under which respondents took the experiment and the norm sensitivity questions, participants were asked to indicate what area they thought the speakers of the audio samples came from. For the TV host condition, they were instructed to base their response on the audio samples rather than on their real-world knowledge about those hosts.

The condition-specific control questions targeted a number of factors we suspected to potentially have an impact on the results of the study. For the condition using first names as labels for the varieties, we asked whether participants knew anyone with those names and whether they liked the names. Similarly, for the TV host condition, respondents were asked if they knew the presenters and whether they liked them. In both these conditions as well as in the condition where 'audio sample 1' and 'audio sample 2' were used as labels, participants were invited to indicate what type of variety they heard (SBD, CBD, dialect or something else). Finally, for the conditions with explicit language

[41] At the start of the experiment, we also included a control mechanism to exclude participants whose computer did not render the sound properly. Respondents were asked to listen to an audio clip containing one word. Only those who typed in the right word when prompted were able to start the study.

labels, a question was included about how beautiful participants deemed the varieties presented by the labels.

3 Results

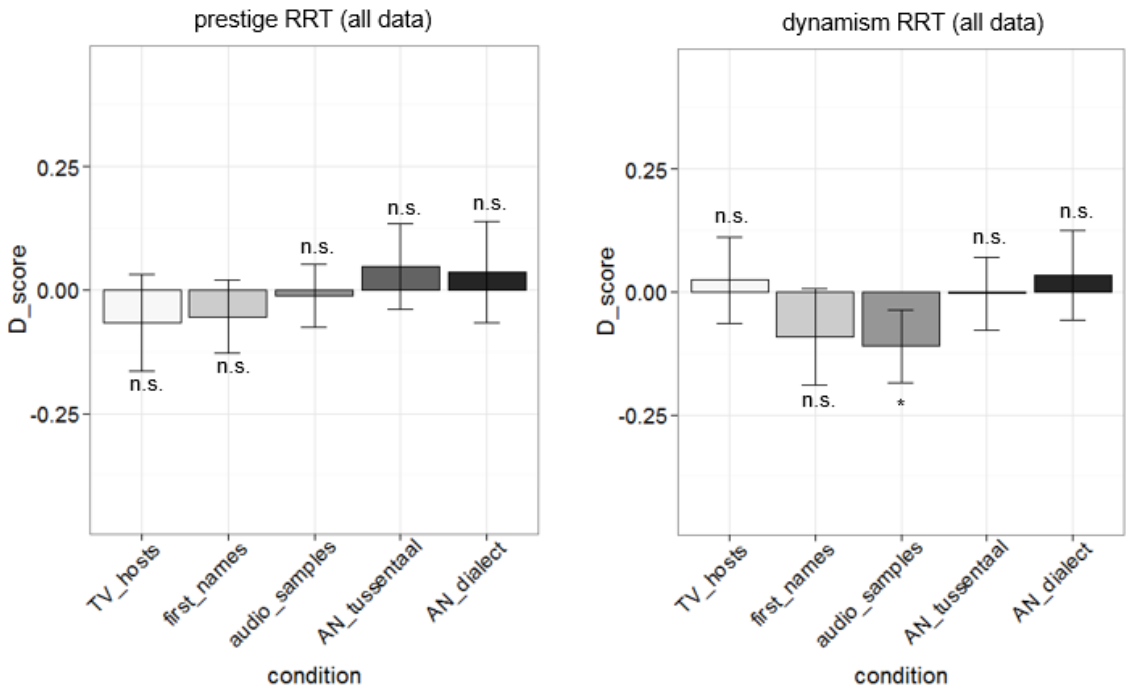
3.1 Implicit attitude measurement

All data for the prestige and dynamism RRTs taken together show no effect for either of the tested beliefs about prestige and dynamism with the exception of the 'focus on audio samples' condition in the dynamism RRT (see Table 16). In the latter condition, participants seem to hold the belief that SBD sounds more dynamic than CBD. That is precisely the opposite of what was hypothesized. We can see this in Figure 14 where the mean D scores are plotted per condition for the prestige and dynamism RRT separately. A positive mean D score indicates a pattern that aligns with our hypotheses (SBD sounds more prestigious than CBD and CBD sounds more dynamic than SBD), while a negative mean score indicates the opposite.

RRT	condition	mean D score	*t*	df	*p*°	*r*
prestige	TV hosts	-0.07	-1.33	64	n.s.	.16
	first names	-0.05	-1.47	51	n.s.	.2
	audio samples	-0.01	-0.34	159	n.s.	.03
	AN-tussentaal	0.05	1.11	62	n.s.	.14
	AN-dialect	0.04	0.73	50	n.s.	.1
dynamism	TV hosts	0.02	0.55	64	n.s.	.07
	first names	-0.09	-1.88	51	n.s.	.25
	audio samples	-0.11	-2.69	159	.02 *	.23
	AN-tussentaal	-0.003	-0.08	62	n.s.	.01
	AN-dialect	0.03	0.76	50	n.s.	.11

° Holm corrected for multiple testing

Table 16 - Overview of the RRT effects (first and second RRT combined)



significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 'n.s.' 1 (Holm corrected)

Figure 14 - D scores for the prestige and dynamism RRTs plotted per condition (all data included). Positive D scores indicate a pattern that is in agreement with the hypotheses (i.e. SBD is more prestigious than CBD and CBD is more dynamic than SBD).

A closer inspection of the data reveals a strong overall order effect depending on whether participants completed the prestige RRT or the dynamism RRT first (see Table 17 for the ANOVA of a regression model showing a significant main effect of RRT order). Scores in the first RRT were generally lower than those in the second RRT (see Figure 15). We suggest this order effect can be interpreted in terms of a training effect. We believe that the RRT procedure may have been so complex for participants to understand and perform that they needed the first few blocks to get the hang of the task they were supposed to carry out. As a result they were slower in the first set of experimental blocks compared to the second set, not because of the beliefs presented in the rules, but because they had to get used to the task. This lead to negative D scores in the first RRT rather than the positive ones we expected. Only in the second RRT were participants well acquainted enough with the procedure for us to measure the effect of following rules that either aligned with

their beliefs or went against those beliefs. This explanation is supported by the data from the second RRTs. When we consider the results for those second RRTs alone, a pattern emerges that comes much closer to the results we hypothesized (see Figure 16). For the prestige RRT, we observe a significant positive effect confirming the hypothesis in three of the five conditions ($M_{\text{TV hosts}}$ = 0.04, $t(32)$ = -0.68, n.s., r = .12; $M_{\text{first names}}$ = -0.04, $t(21)$ = -0.8, n.s., r = .17; $M_{\text{audio sample}}$ = 0.11, $t(82)$ = 2.95, p = 0.02, r = .31; $M_{\text{AN-tussentaal}}$ = 0.14 , $t(28)$ = 2.83, p = .03, r = .47; $M_{\text{AN-dialect}}$ = 0.23, $t(18)$ = 4.11, p < .01, r = .7; all with Holm correction for multiple testing). For the dynamism RRTs we do not find any significant effects, but there is a trend in the data in line with the hypothesized outcome. Although we expected that it would be easier to find evidence for dynamism evaluations for CBD in an implicit measurement of beliefs than in an explicit one, the lack of significant effects for the dynamism dimension, as opposed to the prestige dimension, echoes the suggestion in previous research that positive dynamism evaluations for CBD are hard to detect even when using indirect measures (Grondelaers & Speelman 2013).

model	Predictor	*F* value	*p*	
prestige RRT	RRT order	30.13	< .001	***
	Condition	1.54	.109	n.s.
	RRT order x condition	1.18	.32	n.s.
dynamism RRT	RRT order	36.61	< .001	***
	Condition	2.27	.06	·
	RRT order x condition	1.20	.31	n.s.

model statistics prestige: Adjusted R^2 = .076; F = 4.56, p < .001

model statistics dynamism: Adjusted R^2 = .096; F = 5.61, p < .001

significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 'n.s.' 1

Table 17 - ANOVAs of the linear regression models for prestige and dynamism with the D scores as the dependent variable and RRT order, condition and their interaction as predictors (Chapter 4)

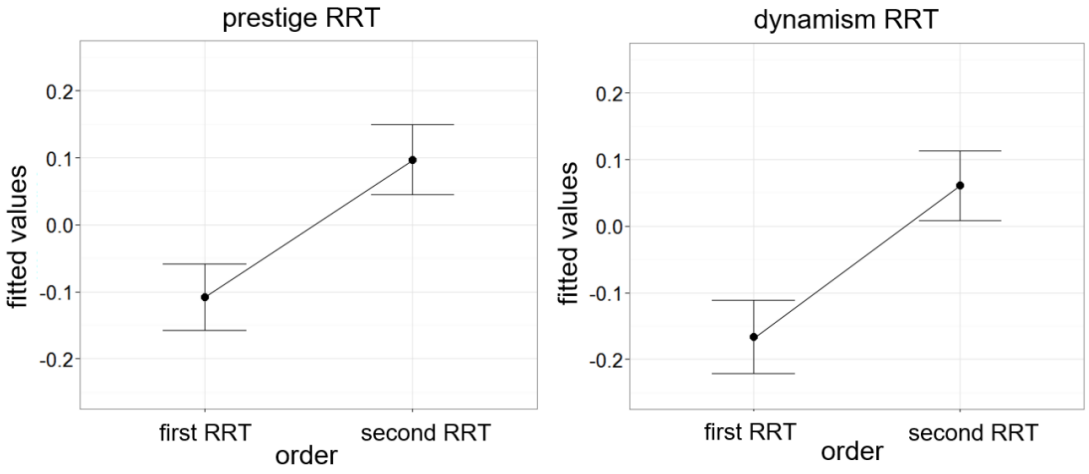
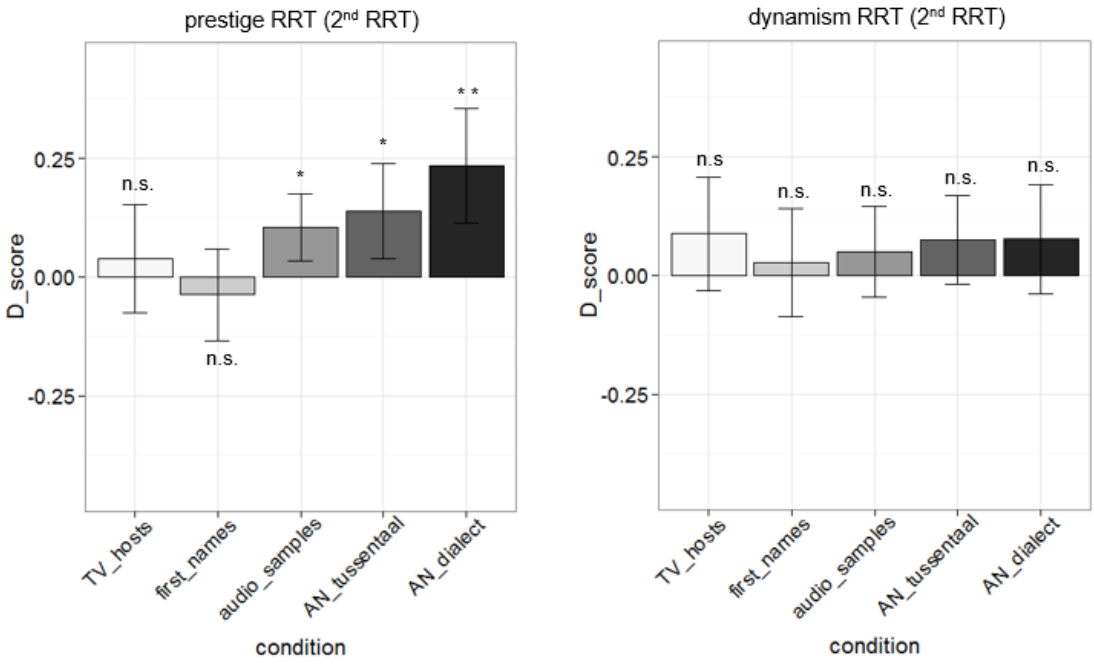


Figure 15 - Effects plot showing the order effect in both the prestige and the dynamism RRT. D scores are lower in the first RRT compared to the second RRT (cf. regression analysis in Table 17).

It is important to stress though that the above interpretation of the order effect is no more than a post hoc explanation and no firm conclusions can be based on this analysis. Further studies would have to be conducted to verify whether similar order effects show up. However, if our interpretation of the order effect is correct, this has serious consequences for the use of the RRT as a measure for the social meaning of language variation. It would mean that a RRT study on language attitudes would have to include a practice RRT that allows participants to get familiar with the method, before the actual RRT could be completed.

Before moving on to the results of the direct rating task, let us briefly consider the impact of the labelling manipulation on the results (of the second RRTs). The five sets of labels in the experiment only affected the prestige RRT ($F(4, 181) = 2.62$, $p = .036$). There we observed significant effects for the 'focus on speech' and the 'focus on audio samples' conditions, but not for the 'focus on speaker' conditions (see Figure 16; note that if we conduct pairwise comparisons between the different conditions, only the first names condition and the *AN-dialect* condition differ significantly, $t(36) = 2.93$, $p = .03$, Tukey corrected for multiple testing). Surprisingly, the labelling manipulation did not affect the dynamism RRTs ($F(4,200) = 0.16$, n.s.). We will come back to this in Section 4 below.



significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 'n.s.' 1 (Holm corrected)

Figure 16 - D scores for the prestige and dynamism RRTs plotted per condition (including only data from participants' second RRT). Positive D scores indicate a pattern that is in agreement with the hypotheses (i.e. SBD is more prestigious than CBD and CBD is more dynamic than SBD).

3.2 Explicit attitude measurement

The results of the explicit measurement will be discussed in three stages starting with an overview of general trends in the results and gradually moving on to a more focused analysis taking into account the different conditions in the experiment. Finally, we will concentrate on a few individual rating scales. Firstly, we want to verify if the overall trends in the data point in the direction of what we hypothesized about our two dimensions, namely that participants associate SBD with prestige and CBD with dynamism (although we were uncertain whether we would be able to find evidence of the latter in the explicit ratings, cf. Section 1.4). At this point, we are aggregating over the five conditions of the experiment. In a second step, we use factor analysis to confirm whether the rating scales indeed represent the two dimension we are interested in (i.e. prestige and dynamism). Here, we distinguish between the five

conditions by running a separate factor analysis for each of the conditions. Finally, we will take a closer look at the differences between the conditions: do the results reflect the cline from focus on speaker to focus on speech that the conditions represent? Additionally, we will also zoom in on a number of specific scales that exhibit interesting patterns.[42]

Figure 17 and Figure 18 summarise the general trends in the direct ratings aggregating over the different conditions in the experiment. The plots clearly show a split between traits expressing prestige/dynamism and those that represent a lack of prestige/dynamism. SBD is associated with traits that indicate prestige (grave, successful, intelligent, smart, rich), while CBD is associated with characteristics indicating a lack of prestige (laidback, insignificant, ignorant, stupid, poor). Adjectives that express dynamism (chill, popular, relaxed, entertaining, trendy) are linked with CBD, while those portraying a less dynamic person (reserved, serious, old-fashioned) are attributed to SBD. A regression analysis confirms this pattern for both dimensions (see Table 18). For each dimension, we built a model with the rating scales of that dimension as the dependent variable. The independent variables were a trait's prestige/dynamism (i.e. whether the trait indicates prestige/dynamism or a lack of it), the experimental condition, and their interaction. The individual participants were included as a random intercept in the model. The two analyses show a significant main effect of a trait's prestige/dynamism, but also a significant interaction of this predictor with condition. As can be seen in Figure 19, the interaction does not counter the

[42] Note that for the analyses and tests in this paragraph, we do not use the factor scores emanating from the factor analysis we carried out. Instead, we work with the rating scales grouped according to the prestige and dynamism dimensions as we defined them at the outset of the study (see Appendix B). As will be discussed, the factor analysis shows the presence of a prestige and dynamism dimension in each condition. However, they differ slightly between the conditions (see Appendix C for a detailed overview of the scales loading onto the different factors in each condition). Hence we deemed it misleading to use the prestige and dynamism factors from the factor analysis given that they are not fully comparable in the five conditions. We decided that maintaining our initial definition of the dimensions when comparing conditions instead would be an acceptable solution (although maybe not ideal), given the evidence for a prestige and dynamism dimension in each of the conditions.

general trend in the aggregated data described above. We will come to the discussion of this interaction with the experimental condition when we take a closer look at the effects of labelling below.

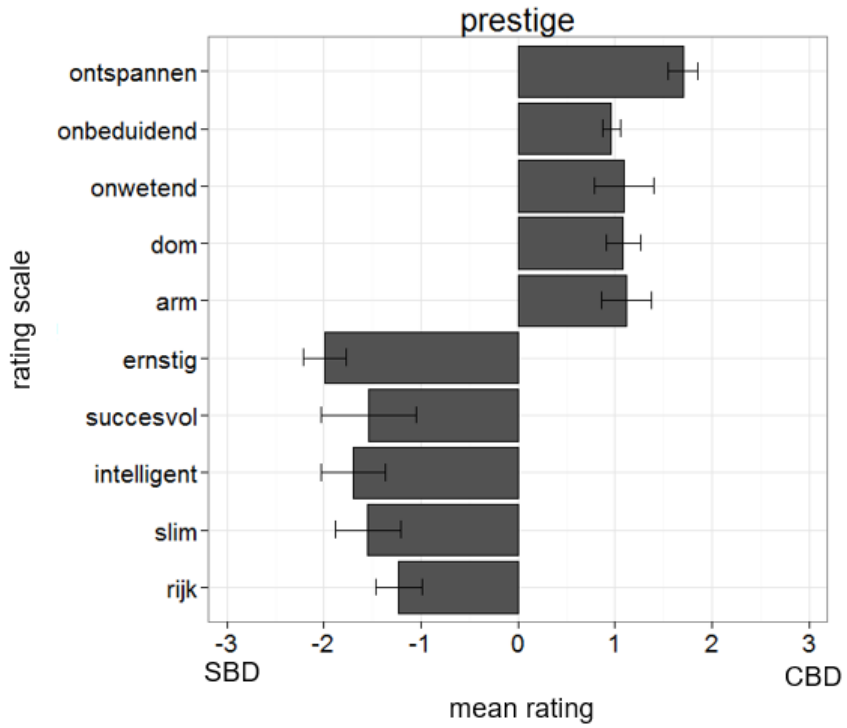


Figure 17 - Mean ratings for the prestige scales in the explicit rating task aggregating over all conditions. Positive scores indicate an association of the traits on the y-axis with CBD, negative scores with SBD.

To find out whether prestige and dynamism were indeed the underlying dimensions in the data, factor analyses were conducted per condition. The best factor solution for each of the conditions is summarised in Table 19[43]. For all five conditions, we find a clear prestige and dynamism dimension as the first two factors. Note though, that the order of the factors and the exact make-up of the prestige and dynamism dimensions differs slightly between the conditions. Additionally, for three of the conditions at the extreme poles of our 'overtness' continuum, namely the TV-hosts condition, the first names condition and the

[43] For a detailed overview of which scales load onto the different factors in each experimental condition, see Appendix C

'AN-dialect' condition, an additional factor was found. For the former two conditions, this was a seriousness/formality dimension. For the latter one, we interpreted the third dimension as 'social insignificance'. Note that these third dimensions are not very robust as they are only based on one or two scales.

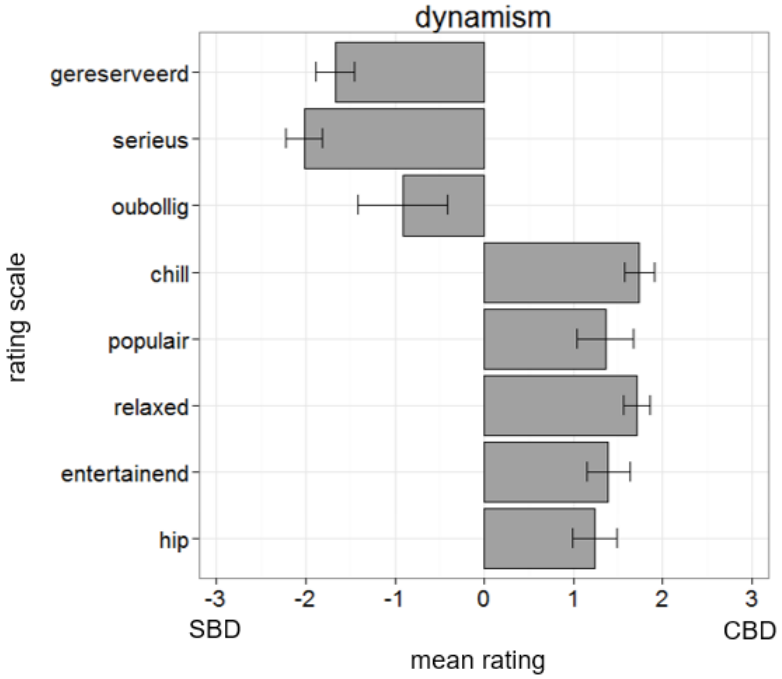


Figure 18 - Mean ratings for the dynamism scales in the explicit rating task aggregating over all conditions. Positive scores indicate an association of the traits on the y-axis with CBD, negative scores with SBD.

Comparing the factor solutions across the different conditions, we would like to point out two observations. Firstly, the condition for which the factor solution appears to be the worst fit, is the condition in which CBD was referred to as 'dialect'. In this condition, the lowest amount of variance is explained by the factor solution and there is a third dimension in addition to prestige and dynamism. One explanation may be that the use of the label 'dialect' evokes social meanings of homeliness and cosiness rather than dynamism. If the experiment had also included rating scales relating to these additional social meanings, that would perhaps have led to a better factor solution in this condition. Secondly, the presence of a seriousness/formality dimension for the two speaker-related conditions is interesting and ties in with

our discussion of situational context in the previous chapter. CBD is associated with informality, while SBD is regarded as more serious.

model	predictor	F value	P	
prestige scales	prestige	6033.28	< .001	***
	condition	5.19	.27	n.s.
	prestige x condition	76.61	< .001	***
dynamism scales	dynamism	4607.55	< .001	***
	condition	4.63	.33	n.s.
	dynamism x condition	22.57	< .001	***

significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 'n.s.' 1

Table 18 - ANOVAs of the linear mixed effects regression models for the prestige and dynamism ratings with participant as a random intercept. The regressors for the prestige model are prestige/dynamism (+prestige/+dynamism vs. -prestige/-dynamism), condition (TV hosts, first names, audio samples, AN-tussentaal, AN-dialect) and their interaction. (Chapter 4)

	Condition				
	TV hosts	first names	audio samples	AN-tussentaal	AN-dialect
factor 1	prestige	dynamism	prestige	dynamism	dynamism
factor 2	dynamism	prestige	dynamism	Prestige	prestige
factor 3	seriousness / formality	seriousness / formality			social insignificance
variation explained	.59	.64	.61	.53	.50

Table 19 - Summary of the factor analysis for each of the five experimental conditions (Chapter 4)

To gain a better understanding of the influence of the labelling manipulation on the outcome of the rating task, let us return to Figure 19 which shows the predicted ratings per condition. We hypothesized that it would be

more likely to find dynamism associations for CBD in the speaker-focused conditions (TV hosts and first names) than in the speech-focused conditions. For the prestige ratings, the opposite was predicted: stronger prestige associations in the conditions where one variety was explicitly labelled as standard Dutch (*AN*). In the upper panel of Figure 19, we see that the hypothesis regarding the prestige dimension is indeed confirmed. Although not all contrasts between the different conditions are statistically significant (see Table 20), overall we can observe a trend that for the conditions with a focus on speech, the association between SBD and prestige, and between CBD and lack of prestige is stronger than in the conditions with a focus on the speaker. In the lower panel of Figure 19, the dynamism ratings do not follow the predicted pattern. In fact, the conditions do not differ significantly from one another with the exception of the 'AN-tussentaal' condition where CBD is rated significantly more dynamic than in the audio samples condition.

model	contrast	prestige / dynamism	significance level
prestige	TV hosts – AN-dialect	pres -	**
	TV hosts – AN-tussentaal	pres -	**
	audio samples – AN-dialect	pres -	*
	audio samples – AN-tussentaal	pres -	*
	TV hosts – first names	pres +	*
	audio samples – AN-tussentaal	pres +	***
	first names – AN-dialect	pres +	**
	TV hosts – AN-tussentaal	pres +	***
	audio samples – AN-dialect	pres +	***
	TV hosts – AN-dialect	pres +	***
dynamism	audio samples – AN-tussentaal	dyn +	**

significance codes: 0 '***' .001 '**' .01 '*' .05 '.' .1 'n.s.' 1

Table 20 - Overview of the significant contrasts in the interaction between prestige/dynamism and experimental condition (with Tuckey correction for multiple comparisons) (Chapter 4)

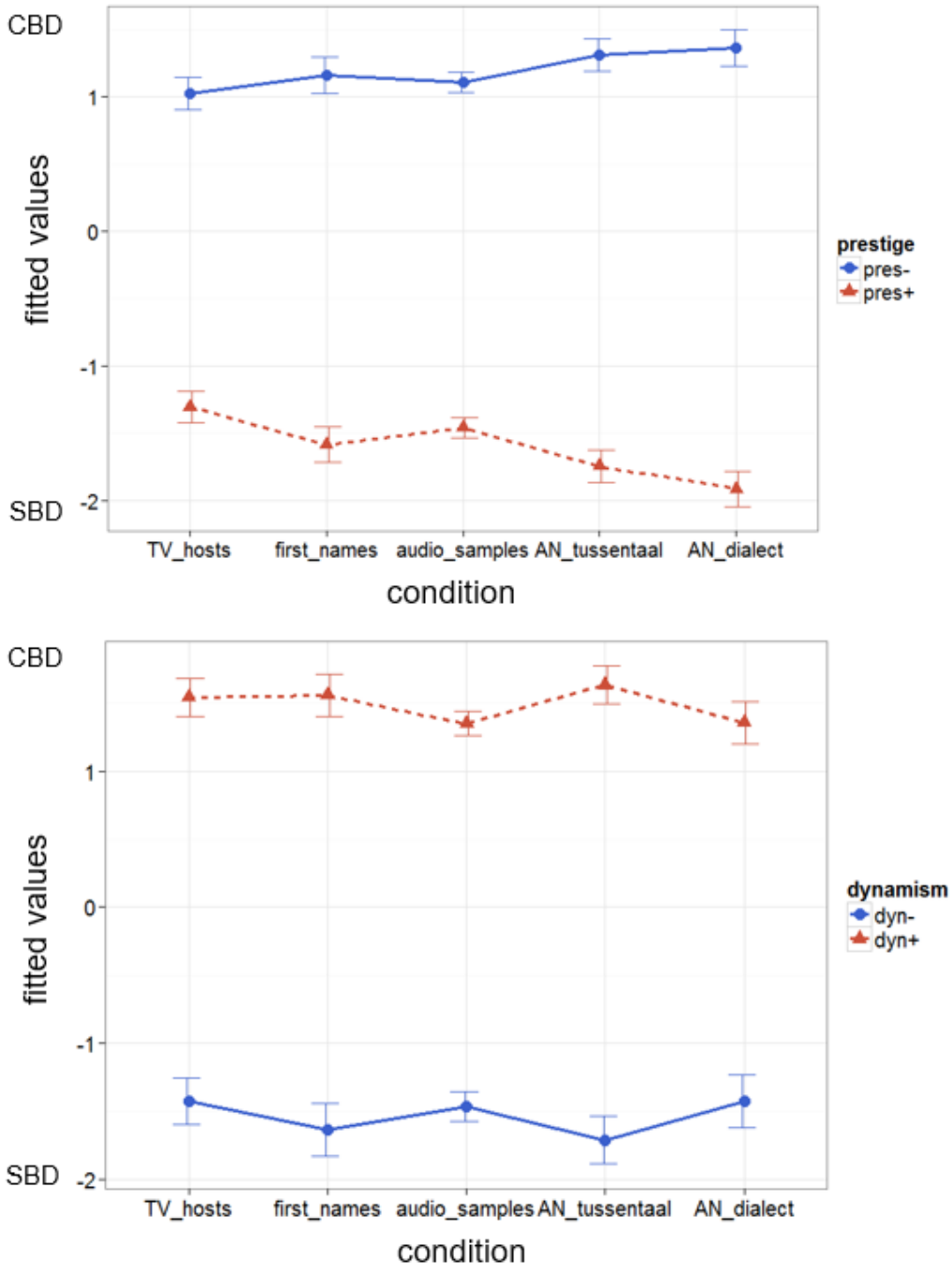


Figure 19 - Interaction plot of the prestige/dynamism effect with condition. Positive scores indicate an association with CBD, negative scores with SBD. Ratings for traits that lack prestige (pres-) or dynamism (dyn-) are plotted in blue. Ratings for traits that indicate prestige (pres+) or dynamism (dyn+) are plotted in red.

As a final step in the analysis of the direct rating task results, let us take a closer look at a number of individual rating scales[44]. This will allow us to get a better insight into the influence of labelling on a more local level in contrast to the aggregate perspective presented above. Firstly, as can be seen in Figure 19, the prestige scores for the TV hosts tend to be the most moderate (i.e. closest to 0). This is particularly clear when we focus on the 'successful' scale (Figure 20). In addition to the explanation for this pattern in terms of focus on speech vs. focus on speaker (i.e. overtness), the specific labels may also have an impact on the ratings. Bart Schols, the TV host assigned to the SBD guise, is rated significantly less successful than the SBD sample in the other 'focus on speech' condition ($M_{\text{Bart Schols}}$ = -0.92, $M_{\text{first names}}$ = -1.71, $t(115) = 3.77$, $p < .001$). A potential explanation for this difference is the fact that both TV hosts are quite successful, as they both present highly popular TV programmes. This interferes with the participants' tendency to assign high successfulness ratings to the SBD guise. A similar explanation can be invoked for the ratings on the 'rich' scale (Figure 21). Although here, there is no significant difference with the other 'focus on speaker' condition, there is a borderline significant difference with the next condition on the continuum ($M_{\text{Bart Schols}}$ = -0.92, $M_{\text{audio sample 1}}$ = -1.27, $t(112)$= -2.16, p = .066 with Holm correction for multiple testing). Again, the specific labels in the TV host condition probably influence participants in their ratings: given that both presenters work for the same public broadcasting channel doing a similar job, they probably earn similar wages. Hence, participants' ratings are more hesitant (i.e. closer to 0) than in other conditions.

[44] For an overview of all individual scales per condition, see Appendix D.

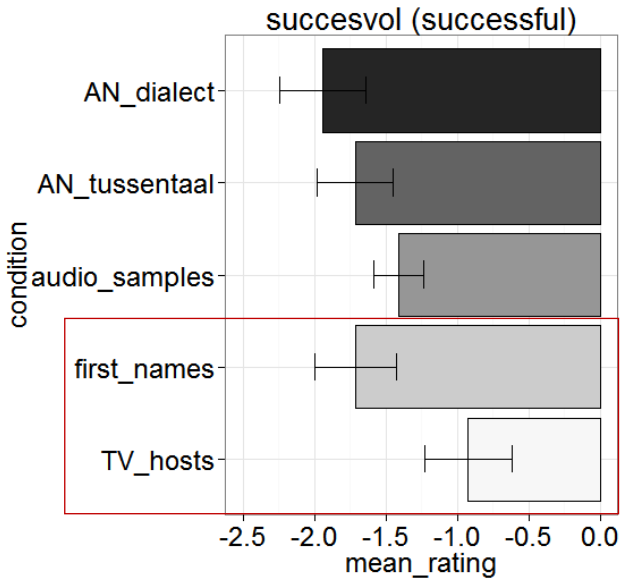


Figure 20 - Mean ratings for the 'successful' scale plotted per condition. Lower scores indicate a stronger association with SBD.

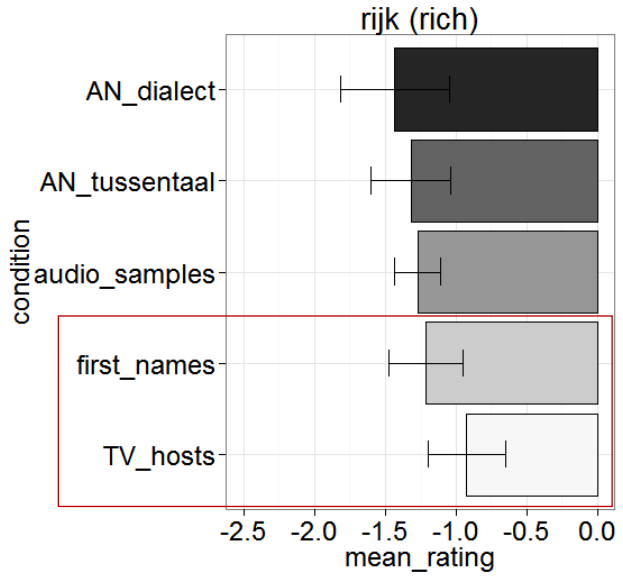


Figure 21 - Mean ratings for the 'rich' scale plotted per condition. Lower scores indicate a stronger association with SBD.

A second pair of rating scales to be discussed is 'trendy' and 'old-fashioned'. Although the overall results indicated no influence of labelling between the 'AN-tussentaal' and 'AN-dialect' condition for the dynamism ratings, we do find significant differences between both conditions for these two crucial dynamism traits. When CBD is referred to as 'tussentaal' it is rated significantly more trendy than when it is referred to as 'dialect' ($M_{tussentaal}$ = 1.41, $M_{dialect}$ = 0.98, $t(103) = 2.05$, p = .04, see Figure 22). For the 'old-fashioned' scale, we can observe that this trait is associated more strongly with SBD when that variety is compared to CBD labelled 'tussentaal' than when it is compared to CBD labelled 'dialect' ($M_{tussentaal}$ = -1.16, $M_{dialect}$ = -0.23, $t(95) = -3.29$, p = .001, see Figure 23). This may be the case because both SBD and dialect are varieties that have been around longer than CBD, which made its appearance more recently (see Section 1.1). Hence, participants did not associate the trait 'old-fashioned' outspokenly with SBD when compared to CBD labelled as 'dialect' given that both varieties are relatively old. These examples of individual scales show that on a lower level of analysis, we do find additional potentially meaningful influence of the different labels used in the five conditions of the experiment.

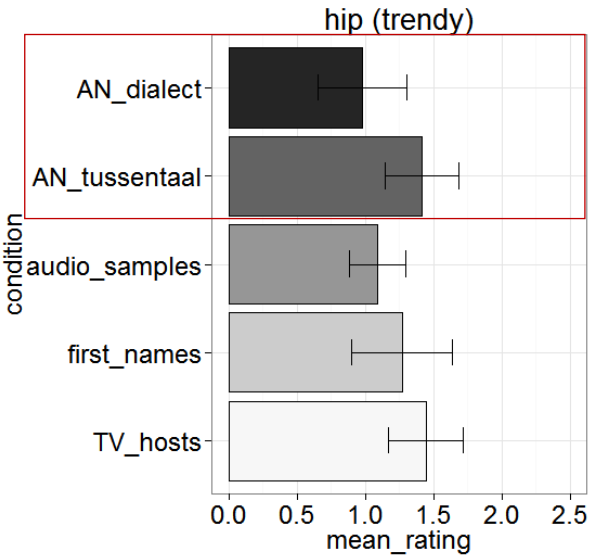


Figure 22 - Mean ratings for the 'trendy' scale plotted per condition. Higher scores indicate a stronger association with CBD.

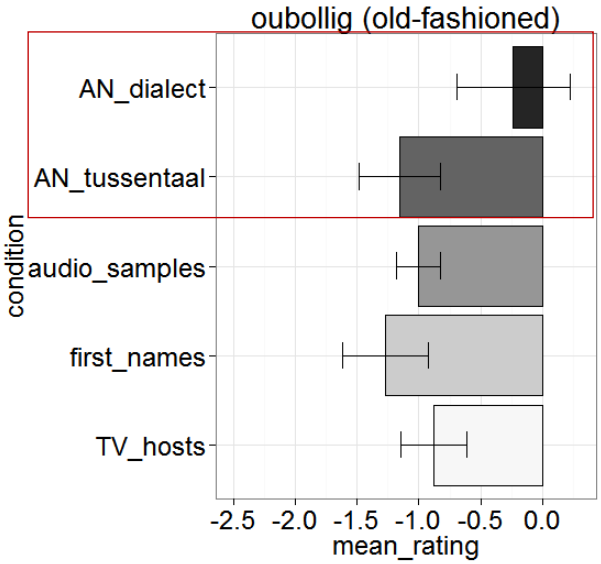


Figure 23 - Mean ratings for the 'old-fashioned' scale plotted per condition. Lower scores indicate a stronger association with SBD.

3.3 Correlation analysis implicit-explicit attitude measurement

Overall, the D scores from the implicit attitude measurement and the aggregate prestige and dynamism scores from the explicit rating task were not correlated. An exception is the dynamism dimension in the 'AN-dialect' condition. Here we find a significant moderate positive correlation between the RRT outcome and the explicit ratings on the dynamism scales ($\rho = .43$, $S = 12597.78$, $p = .002$). A correlation analysis was also conducted on the level of the individual traits comparing specific traits (intelligent, trendy, etc.) across the implicit and explicit measurements. D scores calculated per statement and ratings on the corresponding trait in the direct task were also predominantly not correlated.

Despite this lack of correlation between the two measurements, we can distinguish a similar pattern for the two dimensions across the five conditions in the implicit and explicit attitude measurements. In the explicit measure, traits indicating prestige were most strongly associated with SBD in the two 'focus on speech' conditions, followed by the focus on audio samples condition. In the prestige RRTs (at least in the ones presented as the second RRT) the same three conditions showed a significant association between SBD and

prestige. For the dynamism dimension, there were generally no differences between the conditions for the explicit measurement, which matches the pattern in the dynamism RRTs (that were presented as the second RRT).

3.4 Demographic variables and control questions

As discussed in Section 2.5, a number of control questions were included in the questionnaire following the attitude measurements, complemented by a set of questions gathering demographic information. It was not our goal to study the impact of these factors on the outcome of our attitude measurements, but rather to get an idea of the composition of our sample and detect any potential problems in the design (e.g. participants all know one of the TV hosts, but not the other). Moreover, the sample size is not sufficiently large to test for the effects of all these variables on the outcome of the RRTs and the rating task.

We entered each of the control and demographic variables separately in a linear regression model (per condition) with either the D scores or the explicit ratings as the dependent variable. We will not discuss the results of these analyses in detail, because, as mentioned above, this is not the aim of this study, and also, because the results may not be very reliable given the multiple testing and the relatively small sample sizes. What we will do is briefly summarise the general trends and point to some variables future research may look into in more detail.

Generally, the demographics of the sample did not impact the results and neither did any of the control variables. Two aspects that seemed to be related to small variations in the attitude ratings were geography and norm sensitivity. In some of the conditions there were minor differences between participants from different provinces, yet these were not consistent across conditions and not always straightforwardly interpretable. Regional differences in language attitudes in Flanders have been reported in previous studies as well (e.g. Impe & Speelman 2007; Grondelaers & Van Hout 2011). For the explicit ratings, in most conditions at least one of the norm sensitivity measures correlated with the outcome of the attitude measure. For a summary of these effects, we refer to Appendix E. Although we cannot make any generalizations regarding geography or norm sensitivity based on our data, we

suggest future research may consider further investigating the impact of these factors on language attitudes in Flanders.

4 Discussion

Three aspects of the results reported above need further elaboration. Firstly, the order effect in the RRTs will be discussed. Next, we will come back to the clear dynamism associations with CBD we found in the direct attitude measurement, which was not expected given the reports of previous studies. Finally, the influence of labelling on the results will be looked at in more detail and some potential directions for further research will be suggested.

The RRT results showed a considerable **order effect** which we interpreted as a training effect. Participants needed the first RRT to learn the rather complicated procedure of the measure. A training effect like this has not been reported in previous RRT studies (De Houwer et al. 2015; Koning et al. 2016; Dewitte et al. 2017). Only one of those studies used two consecutive RRTs, yet the order of the RRTs was kept constant in this experiment (Dewitte et al. 2017). A second difference between our design and the design reported in Dewitte et al. (2017) is that their study was administered in laboratory conditions, while our experiment was conducted online (which was motivated by the need to obtain a sufficiently large sample of respondents). It is possible that support from an experiment leader in laboratory conditions helps participants to understand the instructions better. They get the opportunity to ask questions if they are unsure about the procedure they need to follow. Moreover, as discussed in Chapter 1, laboratory conditions allow the researchers to remove sources of distraction. This is not possible when participants take the experiment online in the environment of their choice. It cannot be excluded that these uncontrolled circumstances in which our RRTs were taken have contributed to the order effect we found. So if the order effect can indeed be attributed to participants struggling with the experimental task, one solution would be to take the experiment back to the lab where respondents can receive more support. An alternative solution would be to introduce additional practice trials in the experimental design. This alternative is, of course, not ideal either, as it would make the experiment longer and perhaps

lead to fatigue effects. Whatever solution is chosen, this issue deserves serious consideration in future (linguistic) studies using the RRT.

Returning to our hypotheses (see Section 1.4), we confirmed prestige evaluations for SBD on both the implicit and explicit level. As for CBD, clear **dynamism evaluations** were measured, but only in the direct rating task. The implicit results (for the second RRT) showed only a trend of dynamism associations with CBD. Finding dynamism associations with CBD in the direct measurement was unexpected, based on what previous work has reported (Grondelaers & Speelman 2013) or, alternatively, based on the idea that CBD is a newer variety which is perhaps less entrenched. Yet, the demographics of our sample may explain why we found dynamism evaluations on the explicit level. Participants were young (M_{age} = 20.57, SD = 4.44) and three quarters of the sample was female. As mentioned in Section 1.2, there is some evidence from both production (Plevoets 2008; Van Gijsel et al. 2008; De Caluwe 2009) and perception research (Vandekerckhove & Cuvelier 2007) that CBD is used more often by younger speakers, associated with younger speakers or judged more positively by younger listeners. Plevoets (2008) also reports some gender differences in his study of CBD which align with our results. This may explain why we were able to measure dynamism associations with CBD in our direct measurement. Perhaps, the fact that participants readily judge CBD positively on dynamism traits in an explicit rating task, even in direct comparison with SBD, is also an indication of a change in the stigmatization of CBD in Flanders. Our results confirm that SBD still clearly enjoys prestige evaluations, but rather than perceiving SBD as the best option in every situation, language users may appreciate or even prefer other varieties in specific contexts (cf. Chapter 3). A similar view is also proposed by Lybaert (2014). The finding of positive explicit attitudes towards CBD perhaps also suggests that we should be careful about assuming the perspective that prestige evaluations for SBD following from the top-down imposed standardization process in Flanders are somehow artificial and overshadowing the more authentic and spontaneous dynamism evaluations for CBD which can only be measured indirectly (e.g. Grondelaers & Speelman 2013). Our results indicated the opposite with much stronger direct evidence of dynamism ratings for CBD and only weak associations on the

implicit level, making this an unlikely account, at least for our young sample of participants.

As a final point in this discussion, let us resume the investigation of the **labelling effects** in the experiment. As outlined in Section 1.4, the expectation was that the more overt labels (the 'focus on speech' conditions) would lead to stronger prestige associations than the less overt labels (the 'focus on speaker' conditions). For dynamism, we hypothesized to find the opposite pattern. Additionally, we expected clear prestige and dynamism associations in the 'focus on audio samples' condition, given that the link between the linguistic exemplars in the sound samples and the statements in the RRT/the rating scales was the most easily established.

In the prestige RRT[45], we found RRT effects for the 'focus on audio sample' condition and both the 'focus on speech' conditions. The effect for the former condition was predicted. We hypothesized it would be the easiest to measure associations in this condition as the link between the statements and the sound samples was the most direct. Potential interference of additional associations linked to the labels would be absent as the labels directly refer to the exemplars. The fact that the 'focus on speech' conditions show the expected prestige effects, but the 'focus on speaker' conditions do not, is also in line with what we hypothesized. Using language labels for the varieties may allow the participants to link the sound samples even better to the statements than in the 'focus on audio samples' condition, because the language labels remind them of relevant information about the samples. The explicit ratings on the prestige scales show a similar trend: the most outspoken prestige judgements are to be found in the 'focus on speech' conditions, rather than the 'focus on speaker' conditions. Again, the language labels may have reinforced the social meanings activated by the audio samples leading to stronger judgements. Moreover, in the case of the 'focus on speaker' labels, additional associations with either the first names or the specific TV hosts and their shows may have muddled the picture somewhat leading to significantly weaker prestige associations. In any case, these additional associations are not due to respondents liking or disliking

[45] We will restrict the discussion here to the results of the second RRTs.

the first names or TV hosts presented in the study, as this did not impact the D scores or direct ratings significantly (see Appendix E). Knowing the TV host or people with the first names from the experiments, on the other hand, did influence the outcome of the RRT and rating task in two cases. D scores of participants who knew Bart Schols tend to be slightly higher than those of participants who have only heard of him or do not know him at all[46]. In the explicit rating task, respondents who knew someone called Pieter gave higher prestige ratings to the SBD sample than those who did not know anyone by that name. This supports to some extent our speculation about additional associations with the labels interfering with the associations we aimed to measure in this study. The fact that no effect was found for the 'focus on speaker' conditions in the RRTs could be explained in yet another way: it is possible that participants recognised and categorised the sound samples as different varieties than SBD and CBD, leading to the activation of different associations than those relating to prestige. However, this explanation seems unlikely given that the majority of participants identified the varieties under study the way we intended them (as SBD and CBD). The few respondents that categorised them differently in the control questions did not exhibit any noteworthy deviations in their attitudes.

While the labelling clearly influenced the outcome of both the prestige RRT and the direct ratings on the prestige scales, there was generally no significant difference in outcome between the five conditions for the dynamism dimension. This is rather puzzling as we expected the labelling manipulation in our conditions to be especially relevant for the latter dimension (that is if our conditions do indeed represent an 'overtness' continuum). A potential explanation may be found in the relative nature of our measures, whereby CBD was directly compared to SBD. Perhaps the prestige effects we are measuring are due just as much to CBD lacking prestige as to SBD being associated with it. This may be especially the case when SBD and CBD are explicitly referred to

---

[46] Note that before participants in the TV host condition started the RRTs, they were introduced to the two hosts with a picture and a brief description of who they were and what shows they presented, so all participants would at least have a basic idea of who they were.

with variety labels participants regularly encounter in the media and public debate, where SBD is usually presented as prestigious and respectable. CBD, by contrast, is heavily criticized and, sometimes, its use even causes public indignation (Absilis et al. 2012). These language labels may then activate associations with this heated public debate that mainly centres around prestige. Now, regarding dynamism, it is perhaps the case that SBD is not completely downgraded on the dynamism dimension, creating an asymmetry between the associations of CBD (not prestigious, but dynamic) and SBD (prestigious, but perhaps not undynamic). Hence the use of the 'focus on speech' labels may not activate such polarized associations on the dynamism dimension as it does for the prestige dimension. That could explain why we did not find stronger dynamism associations in the 'focus on speech' conditions (with the exception of the 'AN-tussentaal' condition in the rating task where 'tussentaal' does seem to activate slightly higher dynamism evaluations than in the 'focus on audio samples' condition).

This idea that there is an asymmetry between the two dimensions (CBD being dynamic, but unprestigious, and SBD being prestigious, but perhaps not undynamic) can also suggest that dynamism may not be the ideal dimension to show positive evaluations for CBD. To fully appreciate this point, let us go back to the topic of the previous chapter: situational context. SBD is linked to formal situations in which speakers can both present themselves as professional and knowledgeable (i.e. prestigious), but in which one can also be dynamic. CBD, by contrast, is associated with informal and entertaining situations in which being dynamic and funny is appreciated, but prestige does not play a significant role. This goes to show once again how important the inclusion of contextual factors is in the study of language attitudes. The absence of a situational dimension in this study is a shortcoming which should be addressed in future work.

In addition to subdividing the experimental conditions in terms of overtness, there is another distinction that can be made: in the 'AN-dialect' condition, there is a mismatch between the label 'dialect' and the exemplars respondents are exposed to in the CBD sound sample. As mentioned in Section 1.3, our design does not allow us to come to any firm conclusions regarding the

mechanisms at work behind this mismatch. However, we can see that the results in this condition roughly follow the general pattern. For the prestige dimension, this is not surprising. Previous studies have shown that dialect is perceived as unprestigious, just like CBD (e.g. Vandekerckhove & Cuvelier 2007). So either CBD or dialect would lead to the same results when compared to SBD for prestige. For the dynamism dimension, our results seem to point in the direction of influence of both the exemplars and the label. The exemplar influence is apparent in the fact that the dialect condition follows the same general trends as the other conditions. Evidence for the impact of the label was found in the 'old-fashioned' and 'trendy' rating scales and perhaps in the fact that the factor solutions for this conditions seemed to be the worst fit (see Section 3.2). Nevertheless, we cannot say anything about how participants deal with the conflict between label and exemplars, or whether they even experience a conflict at all. From the control questions for the other conditions (i.e. the 'focus on speaker' and the 'focus on audio samples' conditions), we know that almost all participants categorised CBD as such, which seems to indicate there may have been some cognitive dissonance. Yet, previous research has reported that CBD is not a perceptually clearly delimited variety for many Flemish listeners (Lybaert 2014). As a result, listeners' initial categorisation based on the exemplars may be more malleable and influenced more easily by the categorisation suggested by the label than if CBD had been a perceptually firmly established variety of Dutch.

Even though we cannot come to any firm conclusions regarding the categorisation mechanisms at work here, we do find intriguing influences from our labelling manipulations. This has consequences on a practical level, but also offers perspectives for future research. From a practical point of view, our results add to previous findings that labelling, and by extension categorisation, in perception studies may impact the outcome of an experiment considerably. Regarding future applications of the RRT, we suggest that this method could be explored to investigate which associations are linked to various labels, but on the express condition that we build up a better knowledge of how the RRT functions as a language attitude measure (e.g. get a better grasp of and control over order effects). This would allow us to obtain a better insight in what

different categorisations of language mean to language users. Such explorations may contribute to our understanding of how people conceptualise language variation.

5 Conclusion

In this study we used the RRT as well as a direct rating task to measure the social meanings of SBD and CBD. The study included five conditions in which different labels were used to refer to these two language varieties. A training effect in the RRT aside, we found implicit prestige associations with SBD in the 'focus on audio samples' and 'focus on speech' conditions. For the dynamism dimension, only a non-significant trend of implicit associations with CBD was found. The direct rating task showed a clear pattern of prestige evaluations for SBD and dynamism evaluations for CBD in all conditions. The impact of the labelling manipulation roughly mirrored the one found in the RRTs. Measuring strong explicit dynamism associations with CBD was somewhat unexpected given that previous studies situated these associations on a very 'covert' level. However, we speculate that the demographics of our young and largely female sample may explain this deviation from previous research.

Although some aspects of the RRT as a language attitude measure are not fully understood yet, we suggest it may be an interesting novel tool for studying the social meaning of language variation. More specifically, the RRT could be used in the future to investigate categorisation of language variation from a perceptual perspective, a topic that is still relatively neglected within both the Cognitive Linguistic and the variationist sociolinguistic tradition.

Chapter 5

Conclusions and broader perspectives on cognition and the social meaning of language variation

In this final chapter, the aim is to broaden the perspective of the research reported in this thesis. We will start out by summarising the findings of the three studies discussed in Chapters 2-4 (Section 1). This will bring us to a number of unresolved questions which we deem are crucial to answer and which provide leads for further research (Section 2). The chapter will be concluded by framing our work in a broader theoretical context (Section 3). More specifically, this thesis goes beyond the exploration of novel methods for the study of language attitudes and dovetails with the endeavour of Cognitive Sociolinguistics, a recent theoretical framework that strives to bring together variationist sociolinguistics and Cognitive Linguistics.

1 Overview

In the introductory chapter, it was explained that this research project has a double goal. The main aim is to bring innovation to the quantitative methodology in the field of language attitude research. In addition to that methodological goal, we hoped to make a modest contribution to the description of the language attitude landscape in Flanders. In order to fulfil the methodological goals, we explored the use of two social psychological implicit attitude measures. For the first time in linguistic research, the Personalized Implicit Association Test and the Relational Responding Task were employed to study the social meaning of language variation. For our descriptive purpose, we focused on attitudes towards language at two layers of the stratificational continuum of Dutch in Flanders, namely Standard Belgian Dutch (SBD) and Colloquial Belgian Dutch (CBD). Regarding the latter layer, the studies reported in this thesis featured four different regional varieties of CBD (Antwerp, West-Flemish and Limburgish accented CBD and CBD with a Brabantic flavour), covering a fair part of Flanders.

Chapter 2 reported the successful first exploration of the Personalized Implicit Association Test (P-IAT) as a measure for language attitudes. We studied attitudes towards SBD and two regional varieties of CBD (Antwerp and West-Flemish) in a sample of participants originating from the two corresponding regions. The experiment showed an overall preference for SBD in both groups. This preference for the standard diminished in conditions where participants were presented with their own regional variety in comparison to SBD. We interpreted this as a sign of in-group preference. An alternative explanation was to interpret the effect as a normative reflex: positive attitudes towards SBD increased when compared to an outgroup variety which participants may have experienced as dialectal.

The P-IAT results were put side by side with the outcome of a direct attitude measurement, and with the results of an affective priming experiment involving the same language varieties and conducted by Speelman and colleagues (2013). Generally, the outcomes of the three measurements overlapped, with only minor differences. It was suggested, amongst other explanations, that these points of divergence were due to differences in

categorisation of the attitude objects in AP, the P-IAT and the direct measurement. Whereas AP uses exemplars as stimuli, the direct method relied primarily on category labels and the P-IAT combines both to construe the attitude object.

Chapter 2's discussion of the potential of the P-IAT raised decontextualisation as a point of criticism[47]. The stimuli used in this measure are highly deprived of linguistic or social context. Given the wealth of sociolinguistic and social psychological evidence showcasing the context dependence of (language) attitudes, the study reported in Chapter 3 set out to incorporate situational context in the P-IAT. The measure was used alongside a direct rating task to study attitudes towards SBD and CBD with a Limburgish flavour in formal and informal settings. The situational context was included in the experiment by means of images which were either presented as a backdrop during the categorisation tasks, or in sequences between the P-IAT's blocks. The P-IAT results showed a context effect, but only in interaction with block order and context order. This lead to the tentative conclusion that the inclusion of context in a linguistic P-IAT in this manner is not straightforward. However, successful attempts to include context cues in measures from the IAT paradigm in social psychology suggest that further research into alternative ways of presenting context cues is worth undertaking.

The outcome of the rating task targeting explicit attitudes showed a clear context effect: participants judged SBD favourably in formal contexts and CBD in informal contexts. However, SBD was not downgraded in the informal context, while CBD was perceived negatively in the formal one. These results suggest an asymmetry in the contextualised social meaning of these language varieties.

In addition to two studies exploring the P-IAT, we conducted a study using a second social psychological implicit attitude measure: the Relational Responding Task (RRT). Contrary to the P-IAT which capitalises on measuring whether and to what extent two concepts are associated, the RRT allows the

---

[47] Of course, the P-IAT is not the only experimental method in language attitude research that can be criticized for this reason (see for instance Soukup 2015).

researcher to specify the relation between the two concepts under study by using propositions as stimuli. In Chapter 4 the RRT was employed to measure the social meaning of SBD and CBD with a Brabantic flavour. The hypothesis we set out to test was that SBD is associated with prestige, while CBD is considered to sound dynamic. In order to test this hypothesis, each participant completed two RRTs: one measuring prestige and one measuring dynamism. The experiment consisted of five between participant conditions. Although all participants listened to the same two sound samples representing SBD and CBD prior to completing the RRTs, these samples were labelled differently in each of the five conditions. The labels were chosen so they would represent a continuum of overtness going from focus on speech to focus on speaker. The two 'focus on speech' conditions referred to SBD and CBD using language labels (*AN-tussentaal* and *AN-dialect*) and are situated on the most overt end of the continuum. The 'focus on speaker' conditions, which were considered less overt, labelled the samples using high frequency first names or using the names of two well-known TV hosts. In the final condition the sound samples were labelled 'audio sample 1' and 'audio sample 2'. This last condition featured as a middle category on the continuum between 'focus on speech' and 'focus on speaker'.

Like in Chapters 2 and 3, this study also included a direct rating task, albeit a more elaborate one. In the rating task, participants judged a series of prestige and dynamism traits as being most applicable to either SBD or CBD. The same label manipulation from the RRT was continued in the rating task. Previous studies have situated dynamism associations with CBD at the implicit level and indicated they are difficult to measure using direct methods (e.g. Grondelaers & Speelman 2013). Hence, we hypothesized to find both implicit and explicit prestige evaluations towards SBD, but expected dynamism associations with CBD only to show up in the RRT and not in the rating task.

The overall results from the RRT exhibited a strong order effect depending on whether participants started with the prestige RRT or dynamism RRT. We interpreted this order effect in terms of a training effect. If only the results from the second RRT are considered, prestige evaluations for SBD were recorded in three of the five conditions. With regard to the dynamism

evaluations of CBD, only a trend in that direction was observed. The labelling manipulation did not seem to affect the results in the dynamism RRT, but it did in the prestige RRT. Significant prestige evaluations for SBD were measured in the 'focus on speech' and 'focus on audio samples' conditions, but not in the 'focus on speaker' conditions.

The outcome of the explicit rating task yielded strong evidence of prestige evaluations for SBD and dynamism evaluations for CBD in all conditions. The rather unexpected explicit dynamism attitudes towards CBD were at odds with the reported difficulty of measuring them using direct methods in the literature. We ascribed our findings to the demographics of our test sample which was young and predominantly female. The labelling manipulation in the rating task mirrored the findings for the RRT: virtually no influence on the dynamism ratings, but more pronounced prestige ratings for SBD in the most overt conditions (i.e. the 'focus on speech' conditions). We explained this pattern by suggesting that in the 'focus on speech' and the 'focus on audio samples' conditions, the labels helped participants to make the link between the exemplars presented in the sound samples (played before the RRTs) and the rating task more easily. In addition, we speculated that in the 'focus on speech' conditions, the labels activated similar prestige associations as the exemplars, leading to more outspoken prestige evaluations in both the RRT and the explicit rating task. Furthermore, it was suggested that there may be an asymmetry in the social meanings of SBD and CBD which is reminiscent of the asymmetric results in Chapter 3. On the level of the individual rating scale, we did find evidence that the labelling manipulation had an impact on some dynamism scales as well.

Based on these three studies, what can be concluded about the potential of the P-IAT and the RRT as new tools for linguists to study the social meaning of language variation? We believe these social psychological measures present promising opportunities for linguistics, but on two conditions: an improved technical command of the methods and a better understanding of a number of theoretical phenomena related to (implicit) attitudes.

Firstly, from a technical and practical perspective, the social psychological implicit attitude measures explored in this thesis are not always

easily implemented. They rely on small differences in reaction times, which makes them delicate measures prone to unwanted error variance in the collected data. If participants get distracted or if there are other interfering factors in the design of the experiment, respondents will slow down or speed up for reasons other than the intended experimental manipulation. In the studies reported above, we encountered a number of issues of this kind, such as interactions between block order and context order in Chapter 3, or a strong order effect in the dynamism and prestige RRTs in Chapter 4. We attempted to explain these unexpected effects, but the very fact that they are unexpected suggests that the practical side of these methods is not yet fully under control. Another case in point are the correlations between implicit and explicit attitude measures that were sometimes found, but often absent. The presence or absence of these correlations may have theoretical significance, but they can just as well be ascribed to practical aspects of the experimental design (see Chapter 2).

Secondly, there are multiple theoretical issues relating to (language) attitudes that remain unresolved. Examples are the relationship between implicit and explicit attitudes, or the way listeners categorise language varieties and the impact of these categorisation processes on language evaluation. This leads us to the next paragraph where we will go into these unresolved questions in some more detail. Yet, before doing so, we would like to come back to the view we already articulated in Chapter 1, namely that we can tackle these technical and theoretical lacuna from two directions. On the one hand, we can elaborate and refine (socio)linguistic theory to develop a better understanding of some of the questions discussed below. This will allow us to adapt our methods (novel and traditional) to those new insights and come to better interpretations of the outcome of the measures we use. On the other hand, we firmly believe that one can also take a starting point in methodology to improve theoretical models of language attitudes and language variation and change more generally. Interacting with new methods, like the ones explored in this thesis, can bring theoretical issues to the surface (e.g. awareness vs. automaticity). Given the latter perspective, we are convinced that it is worthwhile to continue the exploration of these new social psychological

attitude measures, alongside the traditional repertory of linguistic attitude measures which have proven their merit over the years.

2 Unanswered questions

As mentioned in Section 1 above, many questions remain to be answered regarding the use of the P-IAT and RRT in linguistics, as well as regarding language attitudes in general. These questions are situated on the **methodological**, descriptive and theoretical level. Above, it was already pointed out that more knowledge is required about the consequences of certain procedural choices in these new social psychological tools. Examples are order effects, the number of tests a participant can take before getting too tired or distracted, or the implementation of context cues. Let us take up the latter issue as an illustration. In Chapter 3, we presented ample evidence from previous research that context plays a crucial role in mediating language attitudes. In that light, it is important to follow up on the study presented in that chapter and further invest in the implementation of context in the P-IAT. Yet, even if we manage to successfully incorporate context cues in a language P-IAT (or perhaps RRT), what we can learn about the influence of contextual information on the social meaning of language variation remains limited. As was suggested in Chapter 2, to fully understand how language attitudes function in context, insights from various methodological approaches are needed. For example, the type of experimental work presented in this thesis could be combined with an interactional approach to language attitudes. A study that can inspire such multimethod explorations is Soukup (2015). Although this study does not include the social psychological implicit attitude measures we are concerned with here, it exemplifies how a quantitative approach to language attitudes building on the speaker evaluation paradigm and a qualitative approach using discourse methods can lead to a richer analysis.

On the **descriptive** side, what is still lacking is a large scale survey of the language attitude landscape in Flanders. In this thesis, we have contributed some pieces to the puzzle, showing for instance firm positive attitudes towards SBD on both the implicit and explicit level. The study in Chapter 4 also confirmed that this variety enjoys strong prestige associations. For CBD, our

studies showed effects of in-group preference and positive evaluation of the variety in informal contexts, but not in formal ones (Chapters 2 and 3). Moreover, we managed to show that it is possible to measure dynamism attitudes towards CBD, even in a direct rating task. Yet, many aspects of the language attitudinal landscape in Flanders remain unexplored. Most studies, like the ones in this thesis, focus on only certain layers of the stratificational continuum of language registers in Flanders, often also working with rather small sample sizes. In addition to that, too little is known about the perception of specific regiolects, sociolects, ethnolects, etc. Take regional variation for example: although this thesis included a range of regionally flavoured varieties of CBD throughout its experiments, a comprehensive and systematic study including a representative sample of regional varieties in Flanders has yet to be carried out. Not only attitudes towards regional variation in language use, but also variation in attitudes amongst participants based on their regional origin needs more thorough investigation. Both the study presented in Chapter 2 and the one discussed in Chapter 4 suggested some regional variation on the side of the listeners. Again this is something that needs to be looked at on a larger scale. A final aspect of language attitudes in Flanders that is in need of more systematic investigation and that we will mention here is the evaluation of specific language variants. As mentioned in Chapter 2, most studies focus on varieties as a whole rather than individual features (with a few exceptions, such as Ghyselen 2009). Of course, large-scale experimental research is challenging and requires a considerable amount of resources. Yet, continuously improving our technical command of methods like the (P-)IAT and the RRT, combined with the possibility to use these techniques in online experiments, should offer opportunities to reach a wide variety of respondents in a time and cost efficient way.

In addition to these descriptive lacuna, a number of questions surfaced in the previous chapters that are situated on a **theoretical** level. We will briefly point out three of them: the relationship between implicit and explicit attitudes and their link with behaviour, the role of categorisation in language perception and the lack of a unified and widely accepted theoretical framework that both offers a detailed model of the cognitive processes involved in language

attitudes, as well as an account of the relationship between perception and production related aspects of language.

The first issue was already hinted at in Chapter 1 when the concepts of implicitness and covertness were introduced. In short, the linguistic interpretation of the over-covert distinction centres around awareness, while the social psychological divide between explicit and implicit attitudes is framed in terms of automaticity. The latter includes awareness, but also other aspects like uncontrollability and speed. The question is whether awareness is the only relevant dimension to language attitudes or whether linguistics could benefit from the multidimensional interpretation of implicitness that is prevalent in social psychology. Additionally, linguistics needs more detailed answers to questions regarding the relationship between covert/implicit and overt/explicit attitudes as well as regarding their link with behaviour (i.e. language variation and change). Let us briefly consider how social psychologists are dealing with these questions and compare that to the current situation in linguistics.

The relationship between implicit and explicit attitudes is a hotly debated topic in social psychology and has been for many years now. In Chapter 4, we discussed a current shift in social psychological theorizing of the processes behind different types of evaluation. While the highly popular dual-process models of attitudes link implicit and explicit attitudes to distinct cognitive processes, some more recent models have suggested that a single process account may be more accurate (Hughes et al. 2011). It goes beyond the scope of this thesis to discuss social psychological views on the relationship between implicit and explicit attitudes in more detail, but the point should at least be made that this field puts a lot of energy in trying to understand the concepts of implicit and explicit attitudes, and a large amount of empirical research is dedicated to validating theoretical models of their relationship and its importance for predicting behaviour.

In linguistics, by contrast, theorizing about the conceptual status and theoretical importance of overt vs. covert language attitudes has been limited and fragmented. Occasionally, scholars make claims about these issues, yet they are rarely built on elaborate theorizing or tested empirically. One example of a linguist who has made suggestions about the role of attitudes for language

variation and change is Labov. Although Labov attributed an important explanatory role to covert language attitudes in the study of language variation and change in his earlier work, his stance seems to have become more anti-subjective in later years (e.g. Labov 1972; ibid. 1984 vs. Labov 2001; see Kristiansen 2010b for a thorough discussion of Labov's changing views on the role of overt and covert attitudes). Following Labov's initial position, other authors sometimes explicitly indicate they believe (covert) language attitudes have an important explanatory role for language variation and change (e.g. Kristiansen & Jørgensen 2005; McKenzie 2010; Kristiansen 2010b; Preston 2013; ibid. 2015). Other language attitude scholars are more hesitant when linking language attitudes and behaviour. A case in point is Garrett (2010: 24-29) who emphasizes the problematic nature of that link based on mixed findings in social psychological attitude research. A few exceptions aside, most empirical linguistic attitude studies do not focus extensively on the link between attitudes and language variation and change. One such exception is the research conducted in Denmark by Kristiansen and colleagues (e.g. Kristiansen & Jørgensen 2005; Grondelaers & Kristiansen 2013) where overt and covert attitudes towards varieties of Danish are linked to language variation and change in Denmark. Similarly, Preston (2015) uses data he collected on language attitudes in Michigan to get a better insight into how a specific language change can occur, in this case the Northern Cities Chain Shift. Nevertheless, more research is needed that empirically tests the relation between overt/explicit and covert/implicit attitudes and between both types of attitudes and language behaviour. Studying these links is interesting in its own right, but it is also a crucial aspect in building a broader theory of language attitudes and their relationship with language production (cf. below). While undertaking this endeavour, linguists should investigate whether their distinction between overt and covert attitudes in terms of awareness is indeed the most relevant approach in the context of language attitude research or whether an interpretation in terms of automaticity, like in social psychology, has added value for their research.

The second theoretical question relates to categorisation processes in language evaluation. In the studies we conducted, this question mainly

surfaced through issues of labelling language varieties (Chapter 4) and through the impact of the type of stimuli used in a task (category labels or exemplars) on its outcome (Chapter 2). If we take a more general perspective, the question can be formulated as: what is the role of individual linguistic features compared to bundles of features or abstractions thereof in language perception? Or even more broadly: how are language varieties represented cognitively? We believe that the study of categorisation processes at work in language perception can shed light on these questions. As has been mentioned in previous chapters, linguistic research has worked with usage-based theories that deal with categorisation processes in terms of prototypicality (e.g. exemplar theory, see Chapter 3). There are examples of studies on the social meaning of language variation that actively engage with these theories (e.g. Kristiansen 2010a; Squires 2013). Let us zoom in on one particular study by way of illustrating the fruitfulness of engaging with these categorisation theories. Kristiansen (2010a) investigates the formation of linguistic stereotypes in children. In an experiment, children of various ages were asked to identify different varieties of Spanish. Based on the outcome of that identification task, combined with the results of an additional questionnaire, the author suggests that children build mental schemas of different language varieties on the basis of the language usage they encounter. These schemas are gradually built up as the child grows older and are primarily based on language features that are socially salient. Hopefully, more work in this line will follow and improve our understanding of categorisation processes in language perception.

A final issue we will discuss here is the lack of a unified theoretical framework in language attitude research. Surely, there are multiple theoretical models of language attitudes, each of them highlighting different angles and valuable in their own right. Yet, many seem to exist alongside one another without much interaction. In what follows, we will illustrate the diversity of these models by briefly listing six examples. Without aiming for a detailed description, we will very succinctly indicate their perspective on language attitudes. A first example is the sociolinguistic monitor as proposed by Labov and colleagues (Labov et al. 2011). This monitor is responsible for determining

the prestige of both perceived and produced language and is conceptualised as being separate from the 'grammar' system.

The sociolinguistic monitor provides a starting point for a second model, which is proposed by Campbell-Kibler (2016). She criticizes the sociolinguistic monitor model for not integrating social cognition with linguistic knowledge, for its focus on socio-economic prestige to the detriment of more diverse and complex forms of social meaning, and for prioritizing conscious processes and relatively neglecting subconscious ones. Instead, she proposes a broader model of meaningful sociolinguistic variation based on insights acquired in third wave sociolinguistics and fields outside linguistics such as cognitive psychology and social cognition research that can inform us about various aspects of human cognition. Campbell-Kibler's model aims to account for both the production of complex sociolinguistic meaning, as well as the perception of that meaning and the processes that link production and perception. In doing so, the model emphasizes that social and linguistic cognition should not be understood as separate, but as an integrated system. Hence, Campbell-Kibler's (2016) model places the study of the social meaning of language variation in a broader context.

Another model of language attitudes can be found in the perceptual dialectological and folk linguistic tradition which investigates how language variation is perceived and mentally represented by non-linguists (e.g. Niedzielski & Preston 2003; Preston 2011; ibid. 2013; ibid. 2015). This model zooms in on the cognitive processes related to language attitudes, also drawing on knowledge available in fields outside linguistics that are concerned with attitudes, perception and human cognition at large. Additionally, attention is devoted to the link between language production, perception, variation and change.

A fourth model of language attitudes can be found in Soukup (2015). This model is more philosophical in nature: it aims to reconcile the differences in ontological status of attitudes in social psychological and constructivist traditions. Regarding the former tradition, Soukup refers to the idea that attitudes are rather fixed entities in memory, while the latter tradition emphasizes the online construction of attitudes in interaction. Tackling this

ontological question allows her to combine the methodologies that are dominant in the two approaches, namely a quantitative approach with roots in social psychology and a qualitative methodology in constructivist work. In order to combine the two visions, Soukup appeals to concepts from the tradition of critical realism.

The penultimate theoretical approach we will mention here can be found in Purschke (2015). This model, referred to as REACT, has a highly similar goal to the one described in Soukup (2015), namely to combine the more traditional view of attitudes as rather stable mental entities and the constructivist perspective that highlights the dynamic nature of attitudes as evaluative practices. In building his REACT framework, Purschke draws on various traditions and fields, ranging from philosophy to social psychology.

While most of the models mentioned above integrate insights from various domains of the social sciences, the final framework we will name stands out for most directly importing a model from one specific field outside linguistics. Pantos (2012) suggests adopting the Associative-Propositional Evaluation (APE) Model from social psychology to account for the cognitive processes relating to language attitudes (see also Pantos & Perkins 2012 and Pantos 2015). APE is a dual process model (see Chapter 4) that links implicit attitudes to associative and explicit attitudes to propositional reasoning. Drawing on the APE model, the authors can theorize the diverging implicit and explicit attitudes they measured in their research on the perception of foreign accented speech.

The examples of theoretical models given above have multiple commonalities, and at the same time diverge on various points. While a thorough comparison of these models would be beyond the scope of this section, we will provide a succinct survey of a number of their characteristics, namely their scope and positioning of language attitudes vis-à-vis other cognitive processes, their take on implicit vs. explicit attitudes, their focus on language (attitudes) from an individual vs. a community perspective, and finally, their interaction with other fields studying human cognition. After comparing the models along these dimensions, we will return to our point

about the lack of a common theoretical framework and briefly indicate how we perceive such a framework with regards to those dimensions.

Firstly, the models differ significantly regarding their scope. Some take a broader perspective aiming for a model of sociolinguistic variation at large, accounting for both language production and language perception like Campbell-Kibler (2016). The perceptual dialectological models may focus more on the perception side of language, but they do not ignore the connection with production either. This wider scope seems to go hand in hand with a view of language in terms of human cognition at large which is present in both of these approaches. Other models postulate processes related to language attitudes as separate from other aspects of language, as is the case for the sociolinguistic monitor. The APE model, the REACT model and the model proposed by Soukup (2015) occupy an intermediary position in this respect. Their focus is more restricted, as they specifically focus on language attitudes, but they do link language attitudes to other forms of cognition.

A second aspect on which the models take different stances is the distinction between implicit and explicit attitudes. Here, two models stand out: the APE model and the sociolinguistic monitor. The former is developed specifically in the light of accounting for the two types of attitudes and their interaction. The latter largely focuses on conscious processes. This is one of the points Campbell-Kibler (2016) criticizes in her revision of the sociolinguistic monitor. Her model acknowledges the importance of both conscious and unconscious processes in language (attitudes). The same goes for the perceptual dialectological model. The model presented in Soukup (2015) does not engage with this issue and neither does the REACT model.

Most models presented above approach language attitudes both from an individual as well as a community level perspective, yet the emphasis often seems to lie on the former. This is especially the case in the APE model. APE considers the cognitive processes responsible for implicit and explicit attitudes as situated in the individual and does not really assume a community level perspective. Similarly, Campbell-Kibler (2016) situates her model on the individual level, yet she explicitly mentions the importance of linking the model

to advances in studies that are concerned with language from a socially situated perspective.

Finally, the models differ with regard to their interaction with fields outside of linguistics. They all seek inspiration in other fields concerned with human cognition to enrich their understanding of linguistic phenomena, although perhaps less so in the case of the sociolinguistic monitor. Yet, each model does so in another way. Pantos & Perkins (2012) directly import a specific model from psychology, while Campbell-Kibler (2016) and the perceptual dialectological model integrate various insights from a wider range of fields including psychology and cognitive science. As already pointed out above, Purschke (2015) and Soukup (2015) cast their net even wider by incorporating philosophical theories and concepts as well.

Remarkably, despite quite some commonalities, these models seem to largely exist alongside one another without much interaction or debate[48]. In other words, what seems to be lacking is an overarching unifying theoretical framework or at least attempts to come to one. Proposing such a model is beyond the scope of this work, but returning to the dimensions along which we compared our six models, we will briefly indicate in which direction an overarching framework can go. In our view, such a framework should be usage-based, include a model for both language production and perception as well as their dialectal relationship, deal with the concept of implicitness and its consequences for language variation and change, conceptualise social and linguistic cognition as a part of human cognition rather than as separate entities, take both an individual as well as a community level perspective on language and in doing so take inspiration from a wide range of other scientific fields. As discussed, some or most of these characteristics are clearly already present in the models mentioned above, yet we believe theorizing about language attitudes will become more solid if it can link up with an overarching theoretical framework of language and cognition. Cognitive Sociolinguistics, as an emerging subfield of Cognitive Linguistics, seems to be a suitable candidate.

---

[48] Campbell-Kibler (2016) is a clear exception here, as the theoretical model proposed in that paper takes the sociolinguistic monitor (Labov et al. 2011) as its starting point.

This theoretical framework aims to bring together the focus on language variation which is central in sociolinguistics, the field from which some of the above models emerge, and the focus on meaning and categorisation which has been one of the foundational elements of Cognitive Linguistics. In the following section, we will elaborate on this suggestion.

3 A broader perspective

The main goal of this thesis has been to introduce and explore new methods for measuring language attitudes while adding to the description of the social meaning of language variation in Flanders. However, it also goes beyond this methodological project. The thesis contributes to congregating language attitude research as part of sociolinguistics, and Cognitive Linguistics, an effort that is central to Cognitive Sociolinguistics (Kristiansen & Dirven 2008; Geeraerts et al. 2010; Pütz et al. 2014). Cognitive Sociolinguistics is an emerging subfield of Cognitive Linguistics that focuses on socially situated language variation. We will briefly introduce Cognitive Sociolinguistics and then indicate how attitudinal research fits into this framework.

Cognitive Sociolinguistics is a developing field of research within Cognitive Linguistics (Kristiansen & Dirven 2008). Its emergence fits in with a relatively new trend in Cognitive Linguistics to turn its attention to language variation, more specifically to intralinguistic variation (Geeraerts & Kristiansen 2015). Cognitive Sociolinguistics strives to bring together variationist sociolinguistics and Cognitive Linguistics. Its aim is to combine the best of both traditions: a tradition of empirical research into language variation from sociolinguistics and a tradition of studying meaning and categorisation from Cognitive Linguistics (Geeraerts et al. 2010; Geeraerts & Kristiansen 2015). To show how language attitude research fits into this picture, we will approach the question from two angles. First, it will be discussed why Cognitive Linguistics needs language attitude research. Subsequently, assuming the reverse perspective, we will suggest what language attitude research has to gain from interaction with Cognitive Linguistics.

To understand the relevance of research on attitudes towards language variation for Cognitive Sociolinguistics, let us first take a step back and consider

more generally why Cognitive Linguistics needs variational linguistics. Firstly, Cognitive Linguistics self-identifies as a usage-based approach to the study of language (e.g. Croft & Cruse 2004; Geeraerts 2005; Dabrowska & Divjak 2015). It is based on the idea that language use is the product of the linguistic system, but that the linguistic system is in turn shaped by language use. A usage-based linguistics studies actual language use, which unavoidably includes language variation (Geeraerts 2005). A usage-based approach to language thus implies the need to study language variation.

Secondly, Cognitive Linguistics is a recontextualising approach to the study of language (Geeraerts 2010a; Geeraerts & Kristiansen 2015). It forms a reaction to a decontextualising trend which started in structuralist linguistics, where the study of the linguistic system (*langue*) took precedence over the study of language usage (*parole*). This continued in the generativist tradition which also focused on the linguistic system and emphasized its universal and psychological character, to the detriment of the study of meaning and function in context and of language as a social phenomenon (see Geeraerts 2010a for a thorough discussion of this idea). Cognitive Linguistics presents a shift in linguistic theorizing as it set out to reposition the study of meaning, function and language use in context at the centre of linguistic research. Cognitive Linguists considers language as collective behaviour of a community that arises from social interaction, rather than as an innate endowment (as has been the dominant view in the generative framework) (Geeraerts & Kristiansen 2014). Language is understood as a way of conceptualising experience and, as a result, is socially and culturally situated. Considering social context in relation to language use is then inextricably linked to the Cognitive Linguistic enterprise of recontextualising linguistic research.

Given its usage-based and recontextualising nature, Cognitive Linguistics needs to include the study of socially stratified language variation. Of course, language variation has been the core business of sociolinguistics since its very beginnings, when it set out to investigate how language variation is structured along social and contextual dimensions. It is clear, then, that this focus on variation and its rigorous empirical study in sociolinguistics can

inspire Cognitive Linguistics to further develop its research on language variation (Geeraerts & Kristiansen 2015).

Now, why should Cognitive Linguistics concern itself specifically with the study of attitudes towards language variation? The answer lies in the study of meaning, which is core business in Cognitive Linguistics (Dabrowska & Divjak 2015). More particularly, there is a specific focus in Cognitive Sociolinguistics to perceive linguistic variation as meaningful[49]. This interest in the meaning of language variation implies the introduction of a subjective dimension into its study of language variation (Geeraerts & Kristiansen 2015). Hence, investigating the perception and evaluation of language variation forms an integral part of the Cognitive Sociolinguistic enterprise.

It may be clear from the above that the study of language variation, and specifically the study of attitudes towards language variation, has a place in the framework of Cognitive Linguistics. Yet, the latter also has contributions to make to language attitude research, namely through its tradition of categorisation research. Categorisation processes play an important part in the study of language attitudes. This became clear at multiple points throughout the research presented in this thesis, especially in Chapters 2 and 4. In Chapter 4, for instance, we saw how labelling can have an impact on language attitudes. Therefore, it is important for language attitude research to have good models of categorisation at its disposal. This is where Cognitive Linguistics comes in. That field has a strong tradition of studying categorisation. As mentioned above, cognitive Linguists consider language as a way of categorising experience (Lakoff 1987; Ramscar & Port 2015). One theory of categorisation Cognitive Linguistic research has heavily relied on is prototype theory (Geeraerts 2010b; Taylor 2015). Originating from psychology (e.g. Rosch 1978), this theory provides an account of how exemplars are structured into categories. It posits that categories can be characterized by the fact that their boundaries are fuzzy rather than clearly delineated, that some members are more typical (i.e. better representatives) of the category than others, that they

---

[49] Note that this view is akin to the third wave sociolinguistic perspective, which conceptualises language variation as a tool for meaning making (Eckert 2012). Speakers use language variation to construct and convey multiple complex identities.

cannot always be defined by a set of necessary and sufficient features and that they exhibit a family resemblance structure with overlaps at the edges of the categories (Geeraerts 2010b). As indicated in Chapter 3, exemplar theory can be understood as an extensional form of prototypicality. Prototype theory has been applied in multiple domains of linguistics (e.g. phonetics, lexical semantics and syntax, Taylor 2015), but it can also be applied to language attitude research. Lects[50] can be thought of as prototypical categories with some category members being more salient than others (Kristiansen 2003). These central members can be language features, but also emblematic speakers of the lect (Kristiansen 2010a) or perhaps also certain situational contexts or other social information (cf. Chapter 3).

Prototypicality is not the only tool for analysis Cognitive Linguistics has in store for language attitude research. The field has also conducted research on cultural models of language variation (e.g. Berthele 2001; Geeraerts 2003; see Holland & Quin 1987 for earlier work on cultural models of language and thought). Cultural models are schematic collections of knowledge about a certain topic shared by a community (Pütz et al. 2014: 10). This branch of research studies the mental models underlying attitudes towards language variation. Geeraerts (2003), for instance, presents two competing ways of thinking about language variation in Western society: the rationalist model, where variation is seen as a threat to emancipation and clear communication, and the romantic model, where variation is perceived as a tool for individual expression. Examples of how this Cognitive Linguistic approach of cultural models can be applied to the empirical study of language attitudes can be found in Berthele (2008) and Berthele (2010). The former study shows how conceptual metaphors (Lakoff & Johnson 1980), another research topic that is central to Cognitive Linguistics, are used to express the rationalist and romantic models of language variation in the context of language policies in Switzerland and the US. Berthele (2010) demonstrates how those two models are combined into a complex model and selectively applied to the varieties of

---

[50] Lect is used in the Cognitive Sociolinguistic tradition as an umbrella term for various types of language varieties: dialects, regiolects, sociolects, ethnolects, idiolects, registers, etc. (e.g. Geeraerts 2005).

Swiss German considered in the study. This complex model allows Berthele to explain why some varieties can be both perceived as prestigious and non-prestigious depending on the perspective that is taken (cf. our results in Chapters 3 and 4 where we reported non-univocal associations with SBD).

With its rich tradition of studying categorisation processes, for instance through prototype theory, but also in the study of cultural models, Cognitive Linguistics offers a theoretical framework which can enrich language attitude research. Within this framework, research can be conducted to answer questions like the ones that surfaced in this thesis. For instance: what is the role of exemplars vs. more abstract (prototypical) categories as represented by different labels in determining attitudes towards language varieties? The results of Chapter 4 seem to suggest that the role of labels is more dominant than that of exemplary speakers. However we do not have much information yet about the impact of typical speakers or linguistic features on language attitudes. This, in turn, raises further questions about how language users come to attach social meanings to those individual language features and how these meanings interact in different lects.

To sum up, this thesis aims to contribute to two research traditions. Our primary goal has been to add to the methodological toolkit of sociolinguistics. Yet, our research also fits in with the emergence of a socio-variationally enriched description of language usage that has been unfolding within Cognitive Linguistics in the last two decades. We hope to have shown that language attitude research and Cognitive Linguistics may benefit from a stronger interaction with each other. Cognitive Sociolinguistics offers them a meeting point for these traditions and a platform to reflect on and study the social meaning of language variation.

References

Abrams, D., & Hogg, M. A. (1987). Language attitudes, frames of reference, and social identity: A Scottish dimension. *Journal of Language and Social Psychology*, *6*(3-4), 201-213.

Absillis, K., Jaspers, J., & Van Hoof, S. (2012). Inleiding. In K. Absilis, J. Jaspers, & S. Van Hoof (Eds.), *De Manke Usurpator. Over Verkavelingsvlaams* (pp. 3-35). Gent: Academia Press.

Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The Handbook of Attitudes* (pp. 173-221). Mahwah, NJ/London: Erlbaum.

Albarracín, D., Johnson B. T., Zanna, M. P., & Kumkale G. T. (2005). Attitudes: Introduction and scope. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The Handbook of Attitudes* (pp. 3-19). Mahwah, NJ/London: Erlbaum.

Albarracín, D., Wang, W., Li, H., & Noguchi, K. (2008). Structure of attitudes. Judgments, memory, and implications for change. In W. D. Crano, & R. Prislin (Eds.), *Attitudes and Attitude Change* (pp. 19-39). New York: Psychology Press.

Auer, P. (2005). Europe's sociolinguistic unity, or: A typology of European dialect/standard constellations. In N. Delbecque, J. Van der Auwera, & D. Geeraerts (Eds.), *Perspectives on Variation: Sociolinguistic, Historical, Comparative* (pp.7-42). Berlin: Mouton De Gruyter.

Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Babel, A. M. (2016). Preface. In A. M. Babel (Ed.), *Awareness and Control in Sociolinguistic Research* (pp. xix-xxii). Cambridge: Cambridge University Press.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society, 39*(4), 437-456.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science, 7*(3), 136-141.

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods, 46*(3), 668-688.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer, & T. K. Srull (Eds.), *The Handbook of Social Cognition* (2nd ed., pp. 1-40). Hillsdale, NJ: Erlbaum.

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record, 60*(3), 527-542.

Berthele, R. (2001). A tool, a bond or a territory. Language ideologies in the US and in Switzerland. *LAUD Paper 533*. Essen: Linguistic Agency of the University of Essen.

Berthele, R. (2008). A nation is a territory with one culture and one language: The role of metaphorical folk models in language policy debates. In G. Kristiansen, & R. Dirven (Eds.), *Cognitive Sociolinguistics. Language Variation, Cultural Models, Social Systems* (pp. 301-331). Berlin/New York: Mouton De Gruyter.

Berthele, R. (2010). Investigations into the folk's mental models of linguistic varieties. In D. Geeraerts, G. Kristiansen, & Y. Peirsman (Eds.), *Advances in Cognitive Sociolinguistics* (pp. 265-290). Berlin/New York: De Gruyter Mouton.

Bishop, H., Coupland, N., & Garrett, P. (2005a). Conceptual accent evaluation: Thirty years of accent prejudice in the UK. *Acta Linguistica Hafniensia: International Journal of Linguistics, 37*(1), 131-154.

Bishop, H., Coupland, N. & Garrett, P. (2005b). Globalisation, advertising and language choice: Shifting values for Welsh and Welshness in Y Drych, 1851-2001. *Multilingua, 24*(4), 343-378.

Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*(5), 828-841.

Bohner, G., Siebler, F., Gonzalez, R., Haye, A., & Schmidt, E. A. (2008). Situational flexibility of in-group-related attitudes: A single category IAT study of people with dual national identity. *Group Processes & Intergroup Relations*, *11*(3), 301-317.

Bosson, J. K., William B. S., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology, 79*(4), 631-43.

Bradac, J. J., Cargile, A. C., & Hallett, J. S. (2001). Language Attitudes: Retrospect, conspect, and prospect. In W. P. Robinson, & H. Giles (Eds.), *The New Handbook of Language and Social Psychology* (pp. 137-155). Chichester: Wiley.

Campbell-Kibler, K. (2007). Accent, (ING), and the social logic of listener perceptions. *American Speech, 82*(1), 32-64.

Campbell-Kibler, K. (2009). The nature of sociolinguistic perception. *Language Variation and Change, 21*(1), 135-156.

Campbell-Kibler, K. (2010). The effect of speaker information on attitudes toward (ING). *Journal of Language and Social Psychology*, 29(2), 214-223.

Campbell-Kibler, K. (2012). The Implicit Association Test and sociolinguistic meaning. *Lingua, 122*(7), 753-763.

Campbell-Kibler, K. (2013). Connecting attitudes and language behavior via implicit sociolinguistic cognition. In T. Kristiansen, & S. Grondelaers (Eds.), *Language (De)standardisations in Late Modern Europe: Experimental Studies* (pp. 307-329). Oslo: Novus.

Campbell-Kibler, K. (2016). Towards a cognitively realistic model of meaningful sociolinguistic variation. In A. M. Babel (Ed.), *Awareness and Control in Sociolinguistic Research* (pp. 123-151). Cambridge: Cambridge University Press.

Campbell-Kibler, K., & McCullough, E. A. (2015). Perceived foreign accent as a predicator of face-voice match. In A. Prikhodkine, & D. R. Preston (Eds.), *Responses to Language Varieties: Variability, Processes and Outcomes* (pp.175-190). Amsterdam/Philadelphia: Benjamins.

Cargile, A. C. (1997). Attitudes toward Chinese-accented speech. An investigation in two contexts. *Journal of Language and Social Psychology, 16*(4), 434-443.

Cargile, A. C., Giles, H., Ryan, E. B., & Bradac, J. J. (1994). Language attitudes as a social process: A conceptual model and new directions. *Language & Communication, 14*(3), 211-236.

Carmichael, K. (2016). Place-linked expectations and listener awareness of regional accents. In A. M. Babel (Ed.), *Awareness and Control in Sociolinguistic Research* (pp. 152-176). Cambridge: Cambridge University Press.

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin, 25*(2), 215-224.

Chu, C. (2013). *Accent-Based Implicit Prejudice: A Novel Application of the Implicit Association Test*. MA dissertation, San Jose State University.

Coupland, N., & Bishop, H. (2007). Ideologised values for British accents. *Journal of Sociolinguistics, 11*(1), 74-93.

Creber, C., & Giles, H. (1983). Social context and language attitudes: The role of formality-informality of the setting. *Language Sciences, 5*(2), 155-161.

Croft, W., & Cruse, D. A. (2004). *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Cuvelier, P. (2007). Standaardnederlands, tussentaal en dialect in Antwerpen. De perceptie van jonge moedertaalsprekers en taalleerders. In D. Sandra, R. Rymenans, P. Cuvelier, & P. Van Petegem (Eds.), *Tussen Taal, Spelling en Onderwijs. Essays bij het Emeritaat van Frans Daems* (pp. 39-58). Gent: Academia Press.

Cvencek, D., Greenwald, A. G., Brown, A. S., Gray, N. S., & Snowden, R. J. (2010). Faking of the implicit association test is statistically detectable and partly correctable. *Basic and Applied Social Psychology, 32*(4), 302-314.

Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2011). Measuring implicit attitudes of 4-year-olds: The preschool Implicit Association Test. *Journal of Experimental Child Psychology, 109*(2), 187-200.

Dabrowska, E., & Divjak, D. (2015). Introduction. In E. Dabrowska, & D. Divjak (Eds.), *Handbook of Cognitive Linguistics* (pp. 1-9). Berlin/Boston: De Gruyter.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*(5), 800-814.

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

De Houwer, J. (2014). A propositional model of implicit evaluation. *Social and Personality Psychology Compass, 8*(7), 342-353.

De Houwer, J., Heider, N., Spruyt, A., Roets, A., & Hughes, S. (2015). The relational responding task: Toward a new implicit measure of beliefs. *Frontiers in Psychology, 6*, article 319.

De Houwer, J., & Moors, A. (2007). How to define and examine the implicitness of implicit measures. In B. Wittenbrink, & N. Schwarz (Eds.), *Implicit Measures of Attitudes: Procedures and Controversies* (pp. 179-194). New York: Guilford.

De Houwer, J., & Moors, A. (2010). Implicit measures: Similarities and differences. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of Implicit Social Cognition: Measurement, Theory and Applications* (pp. 176-196). New York: Guilford.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135*(3), 347-368.

Delarue, S. (2011). Standaardtaal of tussentaal op school? De paradoxale dubbele kloof tussen taalbeleid en taalgebruik. *Studies van de BKL,* 6, 1-20.

Delarue, S., & Lybaert, C. (2016). The discursive construction of teacher identities: Flemish teachers' perceptions of Standard Dutch. *Journal of Germanic Linguistics, 28*(3), 219-265.

De Pascale S. (forthcoming). Cultural models in contact: The case of regional varieties of Italian. In E. Zenner, A. Backus, & E. Winter-Froemel (Eds.), *Cognitive Contact Linguistics*. Berlin: De Gruyter Mouton.

Dewitte, M., De Schryver, M., Heider, N., & De Houwer, J. (2017). The actual and ideal sexual self concept in the context of genital pain using implicit and explicit measures. *The Journal of Sexual Medicine, 14*(5), 702-714.

Drager, K. (2011). Speaker age and vowel perception. *Language and Speech, 54*(1), 99-121.

Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, *12*(4), 453-476.

Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, *41*, 87-100.

Edwards, J. (1982). Language attitudes and their implications amongst English speakers. In E. B. Ryan, & H. Giles (Eds.), *Attitudes towards Language Variation* (pp. 20-33). London: Arnold.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition, 25*(5), 603-637.

Fazio, R. H., & Petty, R. E. (Eds.). (2008). *Attitudes: Their Structure, Function, and Consequences*. New York/Hove: Psychology Press.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*(2), 229-238.

Ferguson, M. J., & Bargh, J. A. (2007). Beyond the attitude object: Automatic attitudes spring from object-centered-contexts. In B. Wittenbrink, & N. Schwarz (Eds.), *Implicit Measures of Attitudes* (pp. 216-246). New York: Guilford.

Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the Implicit Association Test. *Basic and Applied Social Psychology, 27*(4), 307-316.

Foroni, F., & Mayr, U. (2005). The power of a story: New, automatic associations from a single reading of a short scenario. *Psychonomic Bulletin & Review*, *12*(1), 139-144.

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics, 34*(4), 409-438.

Fraeters, F., & Van Avermaet, A. (2010). *Nederlandse Uitspraakoefeningen*. Leuven: Interfacultair Instituut voor Levende Talen, KU Leuven.

Friese, M., Bluemke, M., & Wänke, M. (2007). Predicting voting behavior with implicit attitude measures. *Experimental Psychology, 54*(4), 247-255.

Gallois, C., & Callan, V. J. (1985). Situational influences on perceptions of accented speech. In J. P. Forgas (Ed.), *Language and Social Situations* (pp. 159-173). New York: Springer.

Garrett, P. (2005). Attitude measurements. In U. Ammon, N. Dittmar, K. J. Mattheier, & P. Trudgill (Eds.), *Sociolinguistics: An International Handbook of the Science of Language and Society* (pp. 1251-1260). Berlin: Mouton De Gruyter.

Garrett, P. (2007). Language attitudes. In C. Llamas, L. Mullany, & P. Stockwell (Eds.), *The Routledge Companion to Sociolinguistics* (pp. 116-121). London/New York: Routledge.

Garrett, P. (2010). *Attitudes to Language*. Cambridge: Cambridge University Press.

Garrett, P., Coupland, N., & Williams, A. (2003). *Investigating Language Attitudes: Social Meanings of Dialect, Ethnicity and Performance*. Cardiff: University of Wales Press.

Garrett, P., Williams, A., & Evans, B. (2005). Accessing social meanings : Values of keywords, values in keywords. *Acta Linguistica Hafniensia: International Journal of Linguistics, 37*(1), 37-54.

Gast, A., & Rothermund, K. (2010). When old and frail is not the same: Dissociating category and stimulus effects in four implicit attitude measurement methods. *Quarterly Journal of Experimental Psychology, 63*(3), 479-498.

Gawronski, B., & Bodenhausen, G. V. (2005). Accessibility effects on implicit social cognition: The role of knowledge activation and retrieval

experiences. *Journal of Personality and Social Psychology*, *89*(5), 672-685.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*, 692-731.

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (2nd ed., pp. 283-310). Cambridge: Cambridge University Press.

Gawronski, B., Deutsch, R., & Banse, R. (2011). Response interference tasks as indirect measures of automatic associations. In K. C. Klauer, A. Voss, & C. Stahl (Eds.), *Cognitive Methods in Social Psychology* (pp. 78-123). New York: Guilford.

Gawronski, B., & Hahn, A. (in press). Implicit measures : Procedures, use, and interpretation. In H. Blanton, & G. D. Webster (Eds.), *Foundations of Social Psychological Measurement*. New York: Taylor & Francis.

Gawronski, B., Peters, K. R. & LeBel, E. P. (2008). What makes mental associations personal or extra-personal? Conceptual issues in the methodological debate about implicit attitude measures. *Social and Personality Psychology Compass, 2*(2), 1002-1023.

Gawronski, B., & Sritharan, R. (2010). Determinants and principles of variations in implicit measures. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of Implicit Social Cognition: Measurement, Theory and Applications* (pp. 216-240). New York: Guilford.

Geeraerts, D. (2003). Cultural models of linguistic standardization. In R. Dirven, R. Frank & M. Piitz, (Eds.), *Cognitive Models in Language and Thought. Ideology, Metaphors and Meanings* (pp. 25-68). Berlin: Mouton De Gruyter.

Geeraerts, D. (2005). Lectal variation and empirical data in Cognitive Linguistics. In F. Ruiz de Mendoza Ibáñez, & S. Peña Cervel (Eds.), *Cognitive Linguistics. Internal Dynamics and Interdisciplinary Interactions* (pp. 163–189). Berlin/New York: Mouton De Gruyter.

Geeraerts, D. (2010a). Recontextualizing grammar: Underlying trends in thirty years of Cognitive Linguistics. In E. Tabakowska, M. Choinski, & L. Wiraszka (Eds.), *Cognitive Linguistics in Action: From Theory to Application and Back* (pp.71–102). Berlin/New York: De Gruyter Mouton.

Geeraerts, D. (2010b). *Theories of Lexical Semantics*. Berlin: Mouton De Gruyter.

Geeraerts, D. (2017). Het kegelspel der taal. De naoorlogse evolutie van de Standaardnederlandsen. In G. De Sutter (Ed.), *De Vele Gezichten Van Het Nederlands in Vlaanderen: Een Inleiding Tot De Variatietaalkunde*. Leuven: Acco

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

Geeraerts, D., & Kristiansen, G. (2014). Cognitive linguistics and linguistic variation. In J. Littlemore, & J. Taylor (Eds.), *The Bloomsbury Companion to Cognitive Linguistics* (pp. 202–217). London: Continuum.

Geeraerts, D., & Kristiansen, G. (2015). Variationist linguistics. In E. Dąbrowska, & D. Divjak (Eds.), *Handbook of Cognitive Linguistics* (pp. 365-388). Berlin: De Gruyter Mouton.

Geeraerts, D., Kristiansen, G. & Peirsman, Y. (2010). Introduction. Advances in Cognitive Sociolinguistics. In D. Geeraerts, G., Kristiansen, & Y. Peirsman (Eds.), *Advances in Cognitive Sociolinguistics* (pp. 1-19). Berlin/New York: De Gruyter Mouton.

Geeraerts, D., & Van de Velde, H. (2013). Supra-regional characteristics of colloquial Dutch. In F. Hinskens, & J. Taeldeman (Eds.), *Language and Space. An International Handbook of Linguistic Variation. Volume 3: Dutch* (pp. 532-556). Berlin/Boston: De Gruyter Mouton.

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*. PhD thesis, Ghent University.

Giles, H. (1970). Evaluative reactions to accents. *Educational Review 22*(3), 211-227.

Giles, H., & Ryan, E. B. (1982). Prolegomena for developing a social psychological theory of language attitudes. In E. Ryan, & H. Giles (Eds.), *Attitudes Towards Language Variation: Social and Applied Contexts* (pp. 208-223). London: Arnold.

Glashouwer, K. A., Smulders, F. T. Y., De Jong, P. J., Roefs, A., & Wiers, R. W. H. J. (2013). Measuring automatic associations: Validation of algorithms for the Implicit Association Test (IAT) in a laboratory setting. *Journal of Behavior Therapy and Experimental Psychiatry, 44*(1), 105-113.

Goossens, J. (1970). 'Belgisch Beschaafd Nederlands' en Brabantse expansie. *De Nieuwe Taalgids, 63*, 54-70.

Goossens, J. (2000). De toekomst van het Nederlands in Vlaanderen. *Ons Erfdeel, 43*(1), 3-14.

Govan, C. L., & Williams, K. D. (2004). Changing the affective valence of the stimulus items influences the IAT by re-defining the category labels. *Journal of Experimental Social Psychology, 40*(3), 357-365.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464-1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Attitudes and Social Cognition, 85*(2), 197-216.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji,. M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*(1), 17-41.

Grondelaers, S. (2013). Attitude measurements in the Low Countries. In F. Hinskens, & J. Taeldeman (Eds.), *Language and Space. An*

International Handbook of Linguistic Variation. Volume 3: Dutch (pp. 586-602). Berlin/Boston: De Gruyter Mouton.

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)standardisations in Late Modern Europe: Experimental Studies* (pp. 9-52). Oslo: Novus.

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

Grondelaers, S. & Speelman, D. (2015). A quantitative analysis of qualitative free response data. Paradox or new paradigm? In J. Daems, E. Zenner, K. Heylen, D. Speelman, & H. Cuyckens (Eds.), *Change of Paradigms - New Paradoxes: Recontextualizing Language and Linguistics* (pp. 361-384). Berlin/Boston: De Gruyter Mouton.

Grondelaers, S. & Van Hout, R. (2010). Is Standard Dutch with a regional accent standard or not? Evidence from native speakers' attitudes. *Language Variation and Change, 22*(2), 221-239.

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

Grondelaers, S., & Van Hout, R. (2016). How (in)coherent can standard languages be? A perceptual perspective on co-variation. *Lingua, 172-173*, 62-71.

Grondelaers, S., Van Hout, R., & Speelman, D. (2011). A perceptual typology of standard language situations in the Low Countries. In N. Coupland, & T. Kristiansen (Eds.), *Standard Languages and Language Standards in a Changing Europe* (pp. 199-222). Oslo: Novus.

Grondelaers, S., Van Hout, R., & Steegs, M. (2009). Evaluating regional accent variation in Standard Dutch. *Journal of Language and Social Psychology*, *29*(1), 101-116.

Grondelaers, S., Van Hout, R., & Van Gent, P. (2016). Destandardization is not destandardization. Revising standardness criteria in order to revisit standard language typologies in the Low Countries. *Taal en Tongval, 68*(2), 119-149.

Gschwendner, T., Hofmann, W., & Schmitt, M. (2008). Differential stability. The effects of acute and chronic construct accessibility on the temporal stability of the Implicit Association Test. *Journal of Individual Differences, 29*(2), 70-79.

Hay, J., Drager, K., & Warren, P. (2010). Short-term exposure to one dialect affects processing of another. *Language and Speech, 53*(4), 447-471.

Hay, J., Nolan, A., & Drager, K. (2006a). From fush to feesh: Exemplar priming in speech perception. *Linguistic Review, 23*(3), 351-379.

Hay, J., Warren, P., & Drager, K. (2006b). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics, 34*(4), 458-484.

Heider, N., Spruyt, A., & De Houwer, J. (2016). On the automaticity of relational stimulus processing. *Psychological Research, 81*(1), 1-20.

Hilton, N. H., Rosseel, L., Smidt, E. M., & Coler, M. (2016). Using the IAT to understand the relationship between variant usage patterns and social meaning. Talk presented at *Sociolinguistics Symposium 21*, 15 June 2016, Murcia, Spain.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005a). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*(10), 1369-1385.

Hofmann, W., Gschwendner, T., Nosek, B. A., & Schmitt, M. (2005b). What moderates implicit-explicit consistency? *European Review of Social Psychology, 16*(1), 335-390.

Holland, D., & Quinn, N. (Eds.). (1987). *Cultural Models in Language and Thought*. Cambridge: Cambridge University Press.

Houben, K., & Wiers, R. W. (2006). Assessing implicit alcohol associations with the Implicit Association Test: Fact or artifact? *Addictive Behaviors, 31*(8), 1346-1362.

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorizing in implicit attitude research: Propositional and behavioral alternatives. *The Psychological Record, 61*(3), 465-496.

Hughes, S., Hussey, I., Corrigan, B., Jolie, K., Murphy, C., & Barnes-Holmes, D. (2016). Faking revisited: Exerting strategic control over performance on the Implicit Relational Assessment Procedure. *European Journal of Social Psychology, 46*(5), 632-648.

Huybrechts, G., Decoster, W., Goeleven, A., Lembrechts, D., Manders, E., & Zink, I. (1998). *Articulatie in de Praktijk: Consonanten*. Leuven/Amersfoort: Acco.

Huybrechts, G., Decoster, W., Goeleven, A., Lembrechts, D., Manders, E., & Zink, I. (1999). *Articulatie in de Praktijk: Vocalen en Diftongen*. Leuven/Amersfoort: Acco.

Impe, L. (2006). *Een Attitudineel Mixed Guise-Onderzoek naar Tussentaal in Vlaanderen*. MA dissertation, University of Leuven.

Impe, L. (2010). *Mutual Intelligibility of National and Regional Varieties of Dutch in the Low Countries*. PhD thesis, University of Leuven.

Impe, L., & Speelman, D. (2007). Vlamingen en hun (tussen)taal: Een attittudineel mixed guise-onderzoek. *Handelingen van de Koninklijke Zuid-Nederlandse Maatschappij voor Taal- En Letterkunde en Geschiedenis, 16*, 109-128.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics, 27*(4), 359-384.

Karpinski, A., & Steinman, R. B. (2006). The Single Category Implicit Association Test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91*(1), 16-32.

Katz, D., & Stotland, E. (1959). A preliminary statement to a theory of attitude structure and change. In S. Koch (Ed.), *Psychology: A Study of a Science* (Vol. 3, pp. 423-475). New York: McGraw-Hill.

Koning, I. M., Spruyt, A., Doornwaard, S. M., Turrisi, R., Heider, N., & De Houwer, J. (2016). A different view on parenting: Automatic and explicit

parenting cognitions in adolescents' drinking behavior. *Journal of Substance Use, 22*(1), 1-6.

Koops, C., Gentry, E., & Pantos, A. (2008). The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *University of Pennsylvania Working Papers in Linguistics*, *14*(2), 93-101.

Kristiansen, G. (2003). How to do things with allophones: Linguistic stereotypes as cognitive reference points in social cognition. In R. Dirven, R. Frank, & M. Pütz (Eds.), *Cognitive Models in Language and Thought: Ideologies, Metaphors, and Meanings* (pp. 69–120). Berlin/New York: Mouton De Gruyter.

Kristiansen, G. (2010a). Lectal acquisition and linguistic stereotype formation. In D. Geeraerts, G. Kristiansen, & Y. Peirsman (Eds.), *Advances in Cognitive Sociolinguistics* (pp. 225-264). Berlin/New York: De Gruyter Mouton.

Kristiansen, G., & Dirven, R. (2008). Introduction. Cognitive Sociolinguistics: Rationale, methods and scope. In G. Kristiansen, & R. Dirven (Eds.), *Cognitive Sociolinguistics. Language Variation, Cultural Models, Social Systems* (pp. 1-17). Berlin/New York: Mouton De Gruyter.

Kristiansen, T. (1997) Language attitudes in a Danish cinema. In N. Coupland, & A. Jaworski (Eds.), *Sociolinguistics. A Reader and Coursebook* (pp. 291-305). London: Macmillan.

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

Kristiansen, T. (2010b). Attitudes, ideology and awareness. In R. Wodak, B. Johnston, & P. Kerswill (Eds.), *The Sage Handbook of Sociolinguistics* (pp. 265-278). London: Sage.

Kristiansen, T. (2010c). Conscious and subconscious attitudes towards English influence in the Nordic countries: Evidence for two levels of language ideology. *International Journal of the Sociology of Language, 204*, 59-95.

Kristiansen, T. & Jørgensen, J. N. (2005). Subjective factors in dialect convergence and divergence. In P. Auer, F. Hinskens, & P. Kerswill

(Eds.), *Dialect Change: Convergence and Divergence in European Languages* (pp. 287-302). Cambridge: Cambridge University Press.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (1984). Field methods of the project on linguistic change and variation. In J. Baugh, & J. Sherzer (Eds.), *Language in Use* (pp. 28-53). Englewood Cliffs, NJ: Prentice Hall.

Labov, W. (2001.) *Principles of Linguistic Change. Social Factors* (Vol. 2). Oxford: Blackwell.

Labov, W., Ash, S., Ravindranath, M., Weldon, T., Baranowski, M., & Nagy, N. (2011). Properties of the sociolinguistic monitor. *Journal of Sociolinguistics*, *15*(4), 431-463.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

Lakoff, G. (1990). The invariance hypothesis. Is abstract reason based on image-schemas? *Cognitive Linguistics, 1*(1), 39-74.

Lakoff, G., & Johnson, M. (1980). *Metaphors we Live by*. Chicago/London: The University of Chicago Press.

Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *The Journal of Abnormal and Social Psychology, 60*(1), 44-51.

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the Implicit Association Test : IV. What we know (so far) about the method. In B. Wittenbrink, & N. Schwarz (Eds.), *Implicit Measures of Attitudes* (pp. 59-102). New York: Guilford.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar. Vol. 1.: Theoretical Prerequisites*. Stanford: Stanford University Press.

Lee, R. (2015). Implicit associations with Welsh in two educational contexts. *York Papers in Linguistics, 2*(14), 81-105.

Leinonen, T. (2016). Attitudes, salience, and accommodation-quantity in Finland-Swedish. Talk presented at *ExAPP 2016*, 22 September 2016, Vienna, Austria.

Levon, E. (2007). Sexuality in context: Variation and the sociolinguistic perception of identity. *Language in Society, 36*(4), 533-554.

Liebscher, G., & Dailey-O'Cain, J. (2009). Language attitudes in interaction. *Journal of Sociolinguistics, 13*(2), 195-222.

Llamas, C., & Watt, D. (2014). Scottish , English , British ?: Innovations in attitude measurement. *Language and Linguistics Compass*, *8*(11), 610-617.

Llamas, C., Watt, D., & MacFarlane, A. E. (2016). Estimating the relative sociolinguistic salience of segmental variables in a dialect boundary zone. *Frontiers in Psychology, 7*, article 1163.

Loudermilk, B. C. (2015). Implicit attitudes and the perception of sociolinguistic variation. In A. Prikhodkine, & D. R. Preston (Eds.), *Responses to Language Varieties: Variability, Processes and Outcomes* (pp. 137-156). Amsterdam/Philadelphia: Benjamins.

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology*, *81*(5), 842-855.

Lybaert, C. (2014). *Het Gesproken Nederlands in Vlaanderen. Percepties en Attitudes van een Spraakmakende Generatie*. PhD thesis, Ghent University.

Maddux, W. W., Barden, J., Brewer, M. B., & Petty, R. E. (2005). Saying no to negativity: The effects of context and motivation to control prejudice on automatic evaluative responses. *Journal of Experimental Social Psychology, 41*(1), 19-35.

Maison, D., Greenwald, A. G., & Bruin, R. H. (2004). Predictive validity of the Implicit Association Test in studies of brands, consumer attitudes, and behavior. *Journal of Consumer Psychology, 14*(4), 405-415.

Martin, D. (2015). *IAT: Functions to Use with Data from the Implicit Association Test*. R package version 0.2. http://CRAN.R-project.org/package=IAT.

Marzo, S. (2016). Exploring the social meaning of contemporary urban vernaculars: Perceptions and attitudes about Citélanguage in Flanders. *International Journal of Bilingualism, 20* (5), 501-521.

McKenzie, R. M. (2010). *The Social Psychology of English as a Global Language. Attitudes, Awareness and Identity in the Japanese Context*. Dordrecht: Springer.

McKenzie, R. M. (2015a). UK university students' folk perceptions of spoken variation in English: The role of explicit and implicit attitudes. *International Journal of the Sociology of Language*, *236*, 31-53.

McKenzie, R. M. (2015b). UK university students' perceptions of Japanese, 'local' and other Asian forms of English speech. In M. Yamaguchi (Ed.), *Japanese English in the Era of World Englishes* (pp. 3-12). Kyoto: Kyoto Prefectural University Press.

McKenzie, R. M. (2017). Implicit associations and explicit attitudes towards Northern English and Southern English speech. Invited Talk, *Newcastle University Applied Linguistics Seminar Series*, 7 March 2017, Newcastle, United Kingdom.

Meyerhoff, M. (2006). *Introducing Sociolinguistics*. London/New York: Routledge.

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General, 132*(3), 455-469.

Montgomery, C., & Stoeckle, P. (2013). Geographic information systems and perceptual dialectology: A method for processing draw-a-map data. *Journal of Linguistic Geography, 1*(1), 52-85.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*(2), 297-326.

Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., Van Schie, K., Van Harmelen, A.-L., De Schryver, M., De Winne, J., & Brysbaert, M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4300 Dutch words. *Behavior Research Methods, 45*(1), 169-177.

Nelson, L. R., Signorella, M. L., & Botti, K. G. (2016). Accent, gender, and perceived competence. *Hispanic Journal of Behavioral Sciences, 38*(2), 166-185.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18*(1), 62-85.

Niedzielski, N., & Preston, D. R. (2003). *Folk Linguistics*. Berlin & New York: Mouton De Gruyter.

Nosek, B. A., & Banaji, M. R. (2001). The Go/No-go Association Task. *Social Cognition, 19*(6), 625-664.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002a). E-research: Ethics, security, design, and control in psychological research on the Internet. *Journal of Social Issues, 58*(1), 161-176.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002b). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6*(1), 101-115.

Nosek, B. A., Greenwald, A. G., & Banaji M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality & Social Psychology Bulletin, 31*(2), 166-180.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Automatic Processes in Social Thinking and Behavior* (pp. 265-292). Hove, England: Psychology Press.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality & Social Psychology, 86*(5), 653-667.

Pantos, A. J. (2012). Defining the cognitive mechanisms underlying reactions to foreign accented speech. An experimental approach. *Review of Cognitive Linguistics*, *10*(2), 427-453.

Pantos, A. J. (2015). Applying the Implicit Association Test to language attitudes research. In A. Prikhodkine, & D. R. Preston (Eds.), *Responses to Language Varieties: Variability, Processes and Outcomes* (pp. 117-136). Amsterdam/Philadelphia: Benjamins.

Pantos, A. J., & Perkins, A. W. (2012). Measuring implicit and explicit attitudes toward foreign accented speech. *Journal of Language and Social Psychology, 32*(1), 3-20.

Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality & Social Psychology, 94*(1), 16-31.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality & Social Psychology, 89*(3), 277-293.

Penke, L., Eichstaedt, J., & Asendorpf, J. B. (2006). Single-Attribute Implicit Association Tests (SA-IAT) for the assessment of unipolar constructs. The case of sociosexuality. *Experimental Psychology, 53*(4), 283-291.

Petty, R. E., & Briñol, P. (2009). Implicit ambivalence: A meta-cognitive approach. In R. E. Petty, R. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the New Implicit Measures* (pp. 119-161). New York: Psychology Press.

Petty, R. E., Fazio, R. H., & Briñol, P. (2009). The new implicit measures: An overview. In R. E. Petty, R. H. Fazio, & P. Brinol (Eds.), *Attitudes: Insights from the New Implicit Measures* (pp. 3-18). New York: Psychology Press.

Pharao, N., Maegaard, M., Møller, J. S., & Kristiansen, T. (2014). Indexical meanings of [s+] among Copenhagen youth: Social perception of a phonetic variant in different prosodic contexts. *Language in Society, 43*(1), 1-31.

Pierrehumbert, J. (2001). Stochastic phonology. *GLOT, 5*(6), 1-13.

Plevoets, K. (2008). *Tussen Spreek- en Standaardtaal*. PhD thesis, University of Leuven.

Preston, D. R. (1982). Perceptual dialectology: Mental maps of United States dialects from a Hawaiian perspective. *Working Papers in Linguistics, 14*(2), 5-49. Department of Linguistics, University of Hawaii at Manoa, Honolulu.

Preston, D. R. (1989). *Perceptual Dialectology: Nonlinguists' Views of Areal Linguistics*. Dordrecht: Foris.

Preston, D. R. (1993). The uses of folk linguistics. *International Journal of Applied Linguistics, 3*(3), 181-259.

Preston, D. R. (1996). "Whaddayaknow?": The modes of folk linguistic awareness. *Language Awareness, 5*(1), 40-74.

Preston, D. R. (Ed). (1999). *Handbook of Perceptual Dialectology. Vol. 1.* Amsterdam: Benjamins.

Preston, D. R. (2010). Variation in language regard. In E. Zeigler, P. Gilles, & J. Scharloth (Eds.), *Variatio delectat: Empirische Evidenzen und theoretische Passungen sprachlicher Variation (für Klaus J. Mattheier zum 65. Geburtstag)* (pp. 7-27). Frankfurt am Main: Peter Lang.

Preston, D. R. (2011). The power of language regard - Discrimination, classification, comprehension, and production. *Dialectologia, Special Issue II*, 9-33.

Preston, D. R. (2013). The influence of regard on language variation and change. *Journal of Pragmatics, 52*, 93-104.

Preston, D. R. (2015). Does language regard vary? In A. Prikhodkine, & D. R. Preston (Eds.), *Responses to Language Varieties: Variability, Processes and Outcomes* (pp. 3-36). Amsterdam/Philadelphia: Benjamins.

Preston, D. R. (2016). Whaddayaknow now? In A. M. Babel (Ed.), *Awareness and Control in Sociolinguistic Research* (pp. 177-199). Cambridge: Cambridge University Press.

Preston, D. R., & Niedzielski, N. (2013). Approaches to the study of Language Regard. In T. Kristiansen, & S. Grondelaers (Eds.), *Language (De)standardisations in Late Modern Europe: Experimental Studies* (pp. 287-306). Oslo: Novus.

Price, S., Fluck, M., & Giles, H. (1983). The effect of language of testing on bilingual pre-adolescents' attitudes towards Welsh and varieties of English. *Journal of Multilingual and Multicultural Development, 3*(2-3), 149-161.

Prikhodkine, A. (2015). Attitudes and language detail. Effects of specifying linguistic stimuli. In A. Prikhodkine, & D. R. Preston (Eds.), *Responses to Language Varieties: Variability, Processes and Outcomes* (pp. 219-242). Amsterdam/Philadelphia: Benjamins.

Prikhodkine, A., & Preston, D. R. (Eds.). (2015). *Responses to Language Varieties: Variability, Processes and Outcomes*. Amsterdam/Philadelphia: Benjamins.

Purschke, C. (2015). REACT - A constructivist theoretic framework for attitudes. In A. Prikhodkine, & D. R. Preston (Eds.), *Responses to Language Varieties: Variability, Processes and Outcomes* (pp. 37-53). Amsterdam/Philadelphia: Benjamins.

Pütz, M., Robinson, J. A., & Reif, M. (2014). The emergence of Cognitive Sociolinguistics: An introduction. In M. Pütz, J. A. Robinson, & M. Reif (Eds.), *Cognitive Sociolinguistics. Social and Cultural Variation in Cognition and Language Use* (pp. 1-22). Amsterdam/Philadelphia: Benjamins.

Ramscar, M., & Port, R. (2015). Categorization (without categories). In E. Dabrowska, & D. Divjak (Eds.), *Handbook of Cognitive Linguistics* (pp. 75-99). Berlin: De Gruyter Mouton.

Redinger, D. (2010). *Language Attitudes and Code-switching Behaviour in a Multilingual Educational Context : The Case of Luxembourg*. PhD thesis, The University of York.

Richeson, J. A., & Ambady, N. (2001). Who's in charge? Effects of situational roles on automatic gender bias. *Sex Roles*, *44*(9-10), 493-512.

Rosch, E. (1978). Principles of categorization. In E. Rosch, & B. B. Lloyd (Eds.), *Cognition and Categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.

Rosseel, L., Geeraerts, D., & Speelman, D. (2014). Sociaalpsychologische methodes als nieuwe meettechnieken in taalattitudeonderzoek? Het voorbeeld van de Implicit Association Test. *Handelingen der Koninklijke Zuid-Nederlandse Maatschappij voor Taal- en Letterkunde en Geschiedenis, 68*, 25-39.

Rosseel, L., Speelman, D., & Geeraerts, D. (under review). Measuring language attitudes using the Personalized Implicit Association Test: A case study on regional varieties of Dutch in Belgium.

Rothermund, K., Teige-Mocigemba, S., Gast, A., & Wentura, D. (2009). Minimizing the influence of recoding in the Implicit Association Test:

The Recoding-Free Implicit Association Test (IAT-RF). *Quarterly Journal of Experimental Psychology, 62*(1), 84-98.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education , 33*(4), 511-531.

Rudman, L. A., & Lee, M. R. (2002). Implicit and explicit consequences of exposure to violent and misogynous rap music. *Group Processes & Intergroup Relations, 5*(2), 133-150.

Ryan, E. B., Giles, H., & Sebastian, R. (1982). An integrative perspective for the study of attitudes towards language variation. In E. B. Ryan, & H. Giles (Eds.), *Attitudes Towards Language Variation: Social and Applied Contexts* (pp. 1-19). London: Arnold.

Sarnof, I. (1970). Social attitudes and the resolution of motivational conflict. In M. Jahoda, & N. Warren (Eds.), *Attitudes* (pp. 279-284). Harmondsworth: Penguin.

Scherer, L. D., & Lambert, A. J. (2012). Implicit race bias revisited: On the utility of task context in assessing implicit attitude strength. *Journal of Experimental Social Psychology, 48*(1), 366-370.

Schuurman, I., Schouppe, M., Hoekstra, H., & Van der Wouden, T. (2003). CGN: An annotated corpus of spoken Dutch. In A. Abeille, S. Hansen-Schirra, & H. Uszkoreit (Eds.), *Proceedings of the 4th International Workshop on Linguistically Interpreted Corpora* (pp. 101-108). Budapest.

Schwarz, N. (2007). Attitude construction: Evaluation in context. *Social Cognition, 25*(5), 638-656.

Schwarz, N., & Bohner, G. (2001). The construction of attitudes. In A. Tesser, & N. Schwarz (Eds.), *Blackwell Handbook of Social Psychology: Intraindividual Processes* (pp. 436-457). Malden, MA: Blackwell.

Sherman, J. W., Klauer, K. C., & Allen, T. J. (2010). Mathematical modeling of implicit social cognition: The machine in the ghost. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications* (pp. 156-175). New York: Guilford.

Silverstein, M. (1981). The limits of awareness. *Sociolinguistic Working Paper no. 84*. Austin, TX: Southwest Educational Development Laboratory.

Soukup, B. (2009). *Dialect Use as Interaction Strategy. A Sociolinguistic Study of Contextualization, Speech Perception, and Language Attitudes in Austria*. Vienna: Braumüller.

Soukup, B. (2013a). The measurement of 'language attitudes' - A reappraisal from a constructionist perspective. In T. Kristiansen, & S. Grondelaers (Eds.), *Language (De)standardisations in Late Modern Europe: Experimental Studies* (pp. 251-266). Oslo: Novus.

Soukup, B. (2013b). On matching speaker (dis)guises - Revisiting a methodological tradition. In T. Kristiansen, & S. Grondelaers (Eds.), *Language (De)standardisation in Late Modern Europe: Experimental Studies* (pp. 267-285). Oslo: Novus.

Soukup, B. (2015). Mixing methods in the study of language attitudes. Theory and application. In A. Prikhodkine, & D. R. Preston (Eds.), *Responses to Language Varieties: Variability, Processes and Outcomes* (pp. 55-84). Amsterdam/Philadelphia: Benjamins.

Speelman, D., Spruyt, A., Impe, L., & Geeraerts, D. (2013). Language attitudes revisited: Auditory affective priming. *Journal of Pragmatics, 52*, 83-92.

Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2010). Affect 4.0: A free software package for implementing psychological and psychophysiological experiments. *Experimental Psychology, 57*(1), 36-45.

Spruyt, A., Gast, A., & Moors, A. (2011). The sequential priming paradigm: A primer. In K. C. Klauer, A. Voss, & C. Stahl (Eds.), *Cognitive Methods in Social Psychology* (pp. 48-77). New York: Guilford.

Spruyt, A., Hermans, D., De Houwer, J., & Eelen, P. (2002). On the nature of the affective priming effect: Affective priming of naming responses. *Social Cognition, 20*(3), 227-256.

Spruyt, A., Hermans, D., De Houwer, J., Vandekerckhove, J., & Eelen, P. (2007). On the predictive validity of indirect attitude measures: Prediction of consumer choice behavior on the basis of affective priming

in the picture-picture naming task. *Journal of Experimental Social Psychology, 43*(4), 599-610.

Squires, L. (2013). It don't go both ways: Limited bidirectionality in sociolinguistic perception. *Journal of Sociolinguistics, 17*(2), 200-237.

Staum Casasanto, L. (2008). Does social information influence sentence processing ? *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, 799-804.

Staum Casasanto, L., Grondelaers S., & Van Hout, R. (2015). Got class? Community-shared conceptualizations of social class in evaluative reactions to sociolinguistic variables. In A. Prikhodkine, & D. R. Preston (Eds.), *Responses to Language Varieties: Variability, Processes and Outcomes* (pp. 159-173). Amsterdam/Philadelphia: Benjamins.

Steele, J. R., & Ambady, N. (2006). "Math is hard!" The effect of gender priming on women's attitudes. *Journal of Experimental Social Psychology*, *42*(4), 428-436.

Steffens, M. C. (2004). Is the implicit association test immune to faking? *Experimental Psychology, 51*(3), 165-179.

Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology, 18*(1), 86-100.

Street, R. L. Jr, Bradt, R., & Lee, R. (1984). Evaluative responses to communicators: The effects of speech rate, sex and interaction context. *The Western Journal of Speech Communication, 48*(1), 14-21.

Taylor, J. (2015). Prototype effects in grammar. In E. Dabrowska, & D. Divjak (Eds.), *Handbook of Cognitive Linguistics* (pp. 562-579). Berlin: De Gruyter Mouton.

Teige-Mocigemba, S., Klauer, K. C., & Rothermund, K. (2008). Minimizing method-specific variance in the IAT: A Single Block IAT. *European Journal of Psychological Assessment, 24*(4), 237-245.

Teige-Mocigemba, S., Klauer, K. C., & Sherman, J. W. (2010). A practical guide to Implicit Association Tests and related tests. In B. Gawronski, & K. Payne (Eds.), *Handbook of Implicit Social Cognition: Measurement, Theory and Applications* (pp. 117-139). New York: Guilford.

Thurstone, L. (1931). The measurement of social attitudes. *Journal of Abnormal and Social Psychology, 26*(3), 249-269.

Timmermans, B. (2008). *Klink Klaar. Uitspraak- en Intonatiegids voor het Nederlands*. Leuven: Davidsfonds.

Uhlmann, E., & Swanson, J. (2004). Exposure to violent video games increases automatic aggressiveness. *Journal of Adolescence, 27*(1), 41-52.

Van Bezooijen, R. (2001). Poldernederlands. Hoe kijken vrouwen ertegenaan? *Nederlandse Taalkunde, 6*, 257-271.

Van De Mieroop, D., Zenner, E., & Marzo, S. (2016). Standard and Colloquial Belgian Dutch pronouns of address: A variationist-interactional study of child-directed speech in dinner table interactions. *Folia Linguistica: Acta Societatis Linguisticae Europaeae, 50*(1), 31-64.

Van Gijsel, S., Speelman, D., & Geeraerts, D. (2008). Style shifting in commercials. *Journal of Pragmatics, 40*(2), 205-226.

Van Hoof, S. (2013). *Feiten en Fictie. Een Sociolinguïstische Analyse van het Taalgebruik in Fictieseries op de Vlaamse Openbare Omroep (1977-2012)*. PhD thesis, University of Antwerp.

Van Maele, G. (1984). *Handleiding bij de Uitspraakleer*. Lier: Van In.

Vande Kamp, M. E. (2002). *Auditory Implicit Association Tests*. PhD thesis, University of Washington.

Vandekerckhove, R., & Cuvelier, P. (2007). The perception of exclusion and proximity through the use of standard Dutch, 'tussentaal' and dialect in Flanders. In P. Cuvelier, T. du Plessis, M. Meeuwis, & L. Teck (Eds.), *Multilingualism and Exclusion: Policy, Practice and Prospects* (pp. 241-256). Hatfield, Pretoria: Van Schaik.

Walker, A., García, C., Cortés, Y., & Campbell-Kibler, K. (2014). Comparing social meanings across listener and speaker groups: The indexical field of Spanish /s/. *Language Variation and Change, 26*(2), 169-189.

Watson, K., & Clark, L. (2013). How salient is the nurse~square merger? *English Language and Linguistics, 17*(2), 297-323.

Watt, D., & Llamas, C. (2015). Perception of difference: Socioindexical forms in the Scottish/English border region. Talk presented at *ICLaVE 8*, 27 May 2015, Leipzig, Germany.

Weijnen, A. A. (1946). De grenzen tussen de Oost-Noordbrabantse dialecten onderling. In A. A. Weijnen, J. M. Renders, & J. van Ginneken (Eds.), *Oost-Noordbrabantse Dialectproblemen. Bijdragen en Mededelingen der Dialectencommissie van de Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam 8* (pp. 1-15). Amsterdam.

Weinreich, U., Labov, W., & Herzog, M. I. (1968). Empirical foundations for a theory of linguistic change. In W. F. Lehmann, & Y. Malkiel (Eds.), *Directions for Historical Linguistics* (pp. 95-195). Austin: University of Texas Press.

Wentura, D., & Degner, J. (2010). A practical guide to sequential priming and related tasks. In B. Gawronski, & K. B. Payne (Eds.), *Handbook of Implicit Social Cognition: Measurement, Theory, and Application* (pp. 95-116). New York: Guilford.

Wigboldus, D. H. J., Holland, R. W., & Van Knippenberg, A. (2004). *Single Target Implicit Associations*. Unpublished manuscript.

Willemyns, R. (1979). Bedenkingen bij het taalgedrag van Vlaamse universiteitsstudenten uit Brussel-Halle-Vilvoorde. *Taal en Sociale Integratie, 2*, 141-159.

Willemyns, R. (2007). De-standardization in the Dutch language territory at large. In C. Fandrych, & R. Salverda (Eds.), *Standard, Variation and Language Change in Germanic Languages* (pp. 267-279). Tübingen: Gunther Narr.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology, 72*(2), 262-274.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology, 81*(5), 815-827.

Wittenbrink, B., & Schwarz, N. (2007). *Implicit Measures of Attitudes*. New York/London: Guilford.

Wolfram, W. (2011). Fieldwork methods in language variation. In R. Wodak, B. Johnstone, & P. Kerswill (Eds.), *The Sage Handbook of Sociolinguistics* (pp. 296-311). London: Sage.

Xu, K., Nosek, B., & Greenwald, A. G. (2014). Data from the Race Implicit Association Test on the Project Implicit Demo Website. *Journal of Open Psychology Data, 2*(1), e3.

Zahn, C. J., & Hopper, R. (1985). Measuring language attitudes: The speech evaluation instrument. *Journal of Language and Social Psychology, 4*(2), 113-123.

Appendix A

Direct rating task from the study reported in Chapter 3

1.

a. Hoe sta je tegenover de **accenten** die je in het experiment gehoord hebt, in een **informele situatie** (bv. aan tafel of tijdens een avondje uit met vrienden of familie)?

   Kleur een bolletje: hoe dichter bij een accent, hoe positiever je dat accent vindt.

Limburgs accent	O	O	O	O	O	O	O	neutraal accent

b. Hoe sta je tegenover een **Limburgs accent** (zoals gehoord in het experiment) in een **informele situatie** (bv. aan tafel of tijdens een avondje uit met vrienden of familie)?

negatief	O	O	O	O	O	O	O	positief

c. Hoe sta je tegenover een **neutraal accent** (zoals gehoord in het experiment) in een **informele situatie** (bv. aan tafel of tijdens een avondje uit met vrienden of familie)?

negatief	O	O	O	O	O	O	O	positief

2.

a. Hoe sta je tegenover de **accenten** die je in het experiment gehoord hebt, in een **formele situatie** (bv. tijdens het tv-journaal)?

   Kleur een bolletje: hoe dichter bij een accent, hoe positiever je dat accent vindt.

Limburgs accent	O	O	O	O	O	O	O	neutraal accent

b. Hoe sta je tegenover een **Limburgs accent** (zoals gehoord in het experiment) in een **formele situatie** (bv. tijdens het tv-journaal)?

negatief	O	O	O	O	O	O	O	positief

c. Hoe sta je tegenover een **neutraal accent** (zoals gehoord in het experiment) in een **formele situatie** (bv. tijdens het tv-journaal)?

negatief	O	O	O	O	O	O	positief

Appendix B

Materials RRT and explicit rating task from Chapter 4

Stimuli used in the prestige RRT and dynamism RRT from Chapter 4

	prestige RRT	dynamism RRT
inducer stimuli	*juist, goed, correct, exact, in orde* (synonyms 'true') *mis, onjuist, incorrect, verkeerd, fout* (synonyms 'false')	*juist, goed, correct, exact, in orde* (synonyms 'true') *mis, onjuist, incorrect, verkeerd, fout* (synonyms 'false')
statements**	[SBD] *klinkt rijker dan* [CBD]	[SBD] *klinkt hipper dan* [CBD]
	'[SBD] sounds richer than [CBD]'	'[SBD] sounds trendier than [CBD]'
	[SBD] *klinkt slimmer dan* [CBD]	[SBD] *klinkt entertainender dan* [CBD]
	'[SBD] sounds more clever than [CBD]'	'[SBD] sounds more entertaining than [CBD]'
	[SBD] *klinkt intelligenter dan* [CBD]	[SBD] *klinkt relaxter dan* [CBD]
	'[SBD] sounds more intelligent than [CBD]'	'[SBD] sounds more relaxed than [CBD]'
	[SBD] *klinkt succesvoller dan* [CBD]	[SBD] *klinkt populairder* dan* [CBD]
	'[SBD] sounds more successful than [CBD]'	'[SBD] sounds more popular* than [CBD]'
	[SBD] *klinkt ernstiger dan* [CBD]	[SBD] *klinkt chiller dan* [CBD]
	'[SBD] sounds more serious than [CBD]'	'[SBD] sounds cooler than [CBD]'
	[CBD] *klinkt armer dan* [SBD]	[CBD] *klinkt oubolliger dan* [SBD]
	'[CBD] sounds poorer than [SBD]'	'[CBD] sounds more outdated than [SBD]'
	[CBD] *klinkt dommer dan* [SBD]	[CBD] *klinkt saaier dan* [SBD]
	'[CBD] sounds more stupid than [SBD]'	'[CBD] sounds more boring than [SBD]'
	[CBD] *klinkt onwetender dan* [SBD]	[CBD] *klinkt gestrester dan* [SBD]
	'[CBD] sounds more ignorant than [SBD]'	'[CBD] sounds more stressed out than [SBD]'
	[CBD] *klinkt onbeduidender dan* [SBD]	[CBD] *klinkt minder populair* dan* [SBD]
	'[CBD] sounds more insignificant than [SBD]'	'[CBD] sounds less popular* than [SBD]'
	[CBD] *klinkt zorgelozer dan* [SBD]	[CBD] *klinkt gereserveerder dan* [SBD]
	'[CBD] sounds more carefree than [SBD]'	'[CBD] sounds more reserved than [SBD]'

[SBD] *klinkt armer dan* [CBD]	[SBD] *klinkt oubolliger dan* [CBD]
'[SBD] sounds poorer than [CBD]'	'[SBD] sounds more outdated than [CBD]'
[SBD] *klinkt dommer dan* [CBD]	SBD] *klinkt saaier dan* [CBD]
'[SBD] sounds more stupid than [CBD]'	'[SBD] sounds more boring than [CBD]'
[SBD] *klinkt onwetender dan* [CBD]	[SBD] *klinkt gestrester dan* [CBD]
'[SBD] sounds more ignorant than [CBD]'	'[SBD] sounds more stressed out than [CBD]'
[SBD] *klinkt onbeduidender dan* [CBD]	[SBD] *klinkt minder populair* dan* [CBD]
'[SBD] sounds more insignificant than [CBD]'	'[SBD] sounds less popular* than [CBD]'
[SBD] *klinkt zorgelozer dan* [CBD]	[SBD] *klinkt gereserveerder dan* [CBD]
'[SBD] sounds more carefree than [CBD]'	'[SBD] sounds more reserved than [CBD]'
[CBD] klinkt rijker dan [SBD]	[CBD] *klinkt hipper dan* [SBD]
'[CBD] sounds richer than [SBD]'	'[CBD] sounds trendier than [SBD]'
[CBD] *klinkt slimmer dan* [SBD]	[CBD] *klinkt entertainender dan* [SBD]
'[CBD] sounds more clever than [SBD]'	'[CBD] sounds more entertaining than [SBD]'
[CBD] *klinkt intelligenter dan* [SBD]	[CBD] *klinkt relaxter dan* [SBD]
'[CBD] sounds intelligent than [SBD]'	'[CBD] sounds more relaxed than [SBD]'
[CBD] *klinkt succesvoller dan* [SBD]	[CBD] *klinkt populairder* dan* [SBD]
'[CBD] sounds more successful than [SBD]'	'[CBD] sounds more popular* than [SBD]'
[CBD] *klinkt ernstiger dan* [SBD]	[CBD] *klinkt chiller dan* [SBD]
'[CBD] sounds more grave than [SBD]'	'[CBD] sounds cooler than [SBD]'

* The adjective 'popular' was specified in the instructions to be understood as 'having many friends'.

** In the actual RRTs, the [SBD] and [CBC] slots in the statements were filled with the different labels used for the varieties in the five conditions of the experiment:

- TV hosts: *Bart Schols* and *Otto-Jan Ham*
- first names: *Pieter* and *Jonas* / *Jonas* and *Pieter*
- audio samples: *fragment 1* and *fragment 2*
- language labels (with matching label for CBD): *AN* and *tussentaal*
- language labels (with mismatching label for CBD): *AN* and *dialect*

Traits used in the explicit rating task from Chapter 4

	adjectives
prestige	+ prestige: *rijk, slim, intelligent, succesvol, ernstig* 'rich', 'clever', 'intelligent', 'successful', 'grave'
	- prestige: *arm, dom, ontspannen, onbeduidend, onwetend* 'poor', 'stupid', 'laidback', 'insignificant', 'ignorant'
dynamism	+ dynamism: *chill, populair (heeft veel vrienden), relaxed, entertainend, hip* 'cool', 'popular (i.e. has many friends)', 'relaxed', 'entertaining', 'trendy'
	- dynamism: *gereserveerd, oubollig, serieus* 'reserved', 'outdated', 'serious'

Appendix C

Overview factor analyses per experimental condition

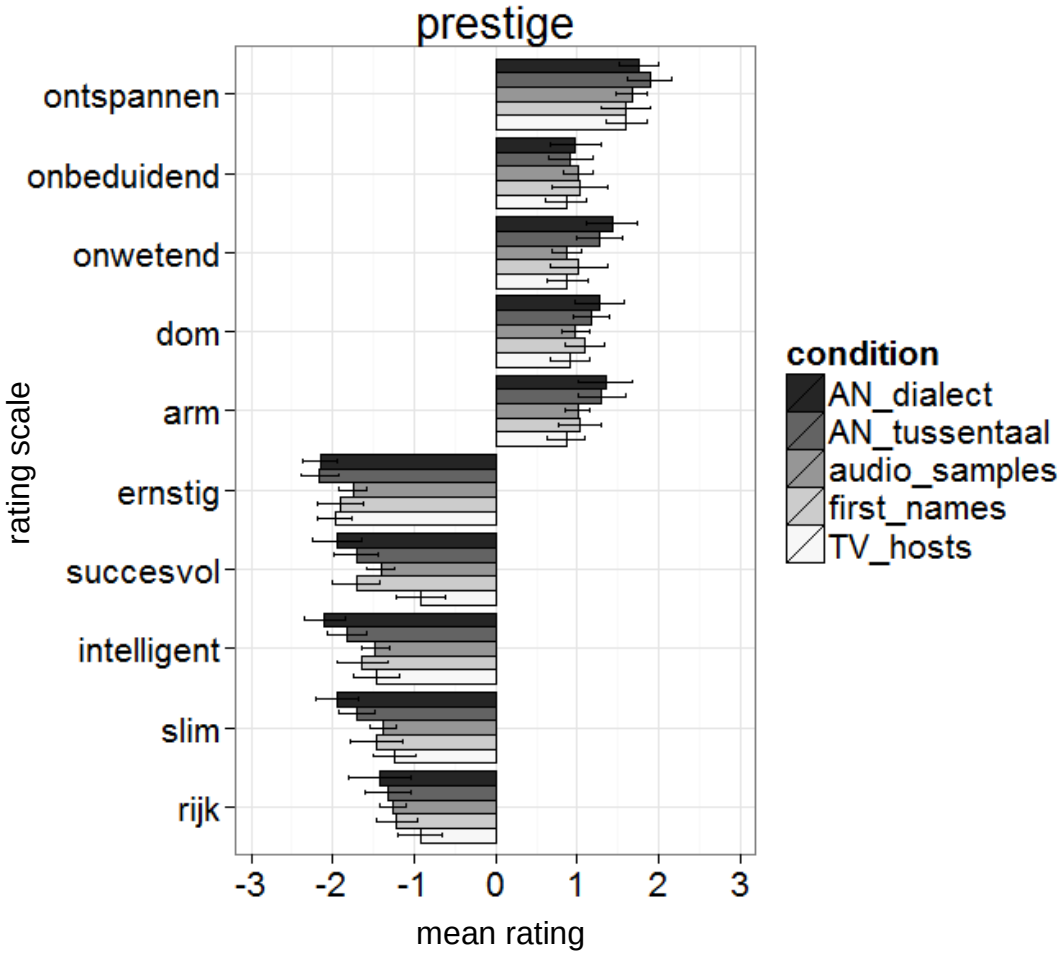
condition	factors	scales	variation explained	rotation
TV hosts	1. prestige	dom – onwetend – arm – slim – rijk – succesvol – onbeduidend - intelligent	0.59	promax
	2. dynamism	hip – entertainend – populair – chill – relaxed – gereserveerd – oubollig		
	3. seriousness/ formality	ernstig – serieus		
first names	1. dynamism	oubollig – hip – chill – populair – entertainend – relaxed – ontspannen – gereserveerd	0.64	promax
	2. prestige	rijk – slim – dom – arm – intelligent – succesvol – onwetend		
	3. seriousness/ formality	ernstig		
audio samples	1. prestige	slim – intelligent – succesvol – dom – onwetend – arm – rijk – onbeduidend	0.61	promax
	2. dynamism	chill – relaxed – hip – populair – entertainend – ontspannen – gereserveerd – oubollig		

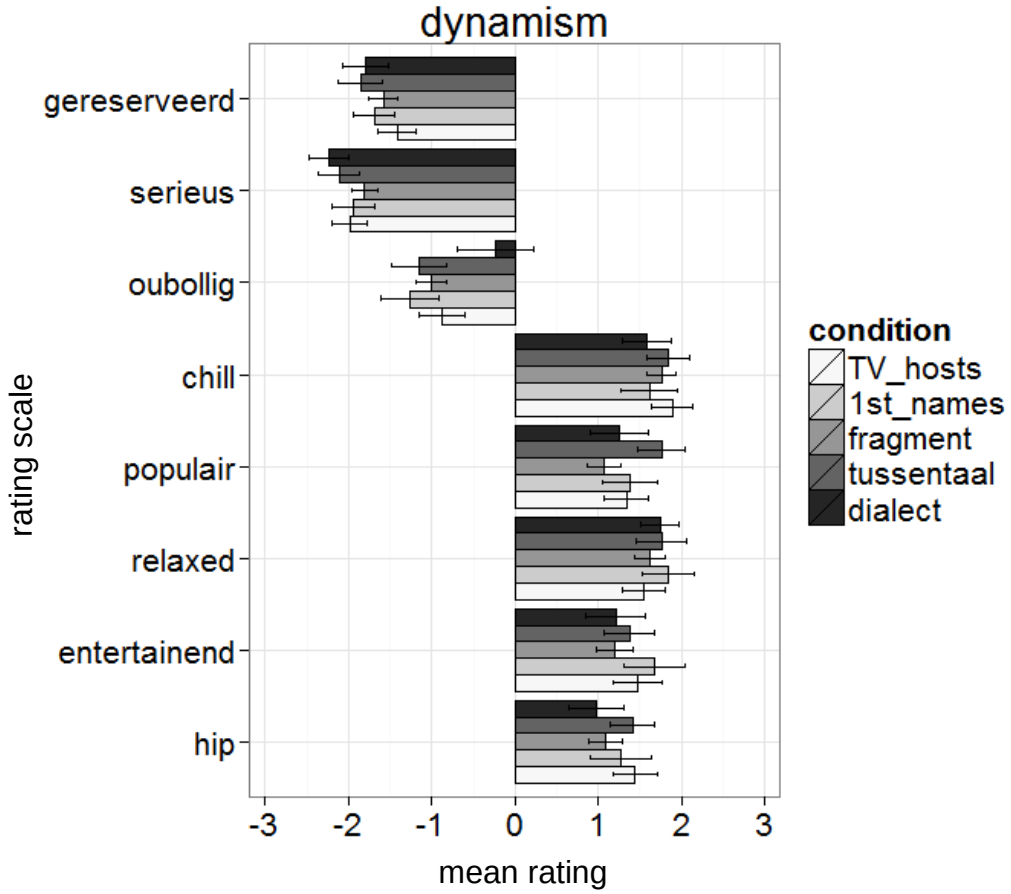
AN-tussentaal	1. dynamism	relaxed – chill – ontspannen – entertainend – hip – populair – oubollig - rijk	0.53	promax
	2. prestige	dom – intelligent – slim – onwetend – arm – onbeduidend – succesvol		
AN-dialect	1. dynamism	chill – populair – ontspannen – relaxed – hip– entertainend	0.50	promax
	2. prestige	slim – intelligent – succesvol – rijk – ernstig – serieus – arm		
	3. social insignificance	onbeduidend		

Scales loading on more than one factor were removed. Note that the third factor for the TV hosts, first names and AN-dialect conditions are not robust given that they are based on one or two scales only.
For translations of the scales, see Appendix B.

Appendix D

Bar plots showing the mean ratings for prestige and dynamism scales grouped per condition




dynamism
gereserveerd
serieus
oubollig
chill
populair
relaxed
entertainend
hip
rating scale
-3
-2
-1
0
1
2
3
mean rating
condition
TV_hosts
1st_names
fragment
tussentaal
dialect

Appendix E

Summary of the effects of the control variables and demographic variables on the D scores of the implicit attitude measurement

Each variable was entered in a linear regression model with the D score for the respective RRT (prestige or dynamism) as the dependent variable and the control/demographic variable as the only regressor. The grey cells in the table indicate that the variable is not applicable in a condition, that there is no effect, or that there was not enough data to reliably test for an effect. For cells with a letter, an effect was found, which is summarised below.

A: The more important it is for a participant to adjust one's language to a specific situation, the less they associate CBD with dynamism.
B: Participants who rate *tussentaal* as more beautiful, associate SBD more strongly with prestige.
C: The more sensitive a participant is for the suitability of a variety in a specific situation, the more they associate SBD with prestige.
D: The more someone likes hearing different accents, the more they tend to associate SBD with prestige.
E: Participants from Antwerp associate prestige with SBD more strongly than those from Limburg and Flemish-Brabant.
F: Participants identifying with the province of Antwerp associate prestige with SBD more strongly than those identifying with Limburg and Flemish -Brabant.
G: Participants identifying with the province of Antwerp associate dynamism with CBD more strongly than those identifying with East-Flanders.
H: Those who like hearing different accents tend to associate SBD more strongly with prestige.
I: Male participants associate prestige more strongly with SBD (vs. CBD) than female participants.
J: Participants wo indicate they do not know where TV host Bart Schols comes from, tend to associate his guise more with dynamism than those who situate him in the centre of Flanders.
K: Respondents who know Bart Schols tend to associate his guise more with prestige than those who have only heard of him or do not know him at all.

L: The more participants disagree with normativity statement 2, the more they associate prestige with SBD.
M: Participants from Antwerp associate prestige more with SBD than their colleagues form Flemish-Brabant.
N: Those who think the SBD speaker originates from the periphery of Flanders find the CBD speaker more prestigious than those who situate him in the centre of Dutch speaking Belgium.
O: Respondents who indicate they do not know where to situate the CBD guise geographically find the SBD speaker more dynamic than those who think the CBD speaker is from the centre of Flanders.

		first names		AN-tussentaal		AN-dialect		TV hosts		audio samples	
		pres	dyn	pres	dyn	pres	dyn	pres	dyn	pres	dyn
Regional origin of the variety	SBD							J		N	
	CBD									O	
Do you like the name [Pieter/Jonas]? / Do you like [Bart Schols/Otto-Jan Ham]?	Pieter/Bart Schols										
	Jonas/Otto-Jan Ham										
Do you know anyone called [Pieter/Jonas] / Do you know [Bart Schols/Otto-Jan Ham]	Pieter/Bart Schols							K			
	Jonas/Otto-Jan Ham										
How would you label the audio samples? (variety label)	SBD										
	CBD										
Normativity measure 1: 'Different types of language are suitable for different types of situations'			A	C							
Normativity measure 2: 'As long as everyone understands what you mean, the type of language you use is not important'								L			
Normativity measure 3: 'I like hearing different accents'				D		H					
Gender						I					
Province of origin of the participants				E				M			
Province participants identify with				F	G						
Beauty ratings	AN										
	tussentaal/dialect			B							

Summary of the effects of the control variables and demographic variables on the ratings in the explicit attitude measurement

Each variable was entered in a linear regression model with the direct prestige and dynamism ratings as the dependent variable and the control/demographic variable as the only regressor. The grey cells in the table indicate that the variable is not applicable in a condition, that there is no effect, or that there was not enough data to reliably test for an effect. For cells with a letter, an effect was found, which is summarised below.

A: Participants who know someone called Pieter give higher prestige ratings to the SBD sample than those who do not know anyone by that name.

B: The more respondents like hearing different accents, the higher they rate CBD for dynamism.

C: Participants from Limburg rate CBD as more dynamic than their counterparts from Antwerp and Flemish-Brabant.

D: East-Flemish respondents find SBD less prestigious than their counterparts from Antwerp, Limburg and Flemish-Brabant.

E: East-Flemish respondents find CBD less dynamic than their counterparts from Antwerp.

F: Participants from East-Flanders find SBD less prestigious than their counterparts from Antwerp.

G: Participants who identify as East-Flemish find CBD less dynamic than their counterparts from Antwerp.

H: The more beautiful participants rate SBD, the more prestige they associate with that variety.

I: People who find dialect beautiful, tend to judge CBD as more dynamic.

J: Respondents from Limburg rate SBD as more prestigious than those from Antwerp.

K: Participants who believe the SBD sample to be CBD rather than SBD, give lower dynamism ratings to CBD.

L: The more important someone finds it to adjust one's language variety to a specific context, the more prestige they associate with SBD.

M: The more important someone finds it to adjust one's language variety to a specific context, the more dynamic they judge CBD.

N: The stronger participants disagree with the statement that language variety is not important as long as one is understood, the more prestigious they find SBD.

O: Respondents from Limburg and Flemish-Brabant find CBD more dynamic than their colleagues from Antwerp.

P: Respondents identifying with the provinces of Limburg, West-Flanders and Flemish-Brabant find CBD more dynamic than their colleagues from Antwerp.

		first names		AN-tussentaal		AN-dialect		TV hosts		audio samples	
		pres	dyn	pres	dyn	pres	dyn	pres	dyn	pres	dyn
Regional origin of the variety	SBD										
	CBD										
Do you like the name [Pieter/Jonas]? / Do you like [Bart Schols/Otto-Jan Ham]?	Pieter/Bart Schols										
	Jonas/Otto-Jan Ham										
Do you know anyone called [Pieter/Jonas] / Do you know [Bart Schols/Otto-Jan Ham]	Pieter/Bart Schols	A									
	Jonas/Otto-Jan Ham										
How would you label the audio samples? (variety label)	SBD										
	CBD										
Normativity measure 1: 'Different types of language are suitable for different types of situations'										L	M
Normativity measure 2: 'As long as everyone understands what you mean, the type of language you use is not important'										N	
Normativity measure 3: 'I like hearing different accents'			B								
Gender											
Province of origin of the participants			C	D	E				J		O
Province participants identify with				F	G						P
Beauty ratings	AN					H					
	tussentaal/dialect						I				