Measuring language attitudes using the Personalized Implicit Association Test: A case study on regional varieties of Dutch in Belgium

Laura Rosseel,[1,2]* Dirk Speelman,[1] and Dirk Geeraerts[1]

[1] *University of Leuven, Leuven, Belgium*
[2] *FWO, Research Foundation Flanders, Brussels, Belgium*

This paper introduces the Personalized Implicit Association Test (P-IAT, Olson & Fazio, 2004) as a novel measure for language attitudes. Tying in with sociolinguists' renewed interest in social psychological attitude measures (e.g., Campbell-Kibler, 2012; Pantos & Perkins, 2012; Speelman, Spruyt, Impe & Geeraerts, 2013), the study uses the P-IAT to measure associations with regional varieties of Belgian Dutch and compares the results to an explicit measurement, as well as the results from an experiment using auditory affective priming, another reaction time based attitude measure developed in social psychology, reported in Speelman et al. (2013). Results from both implicit measures show a strong preference for the standard variety of Belgian Dutch over out-group regional varieties, as well as in-group preferences for participants' own variety over other regional varieties. However, results do not entirely coincide. The paper concludes by discussing the benefits and potential demerits of the P-IAT as a measure of language attitudes.

1. INTRODUCTION

In this paper we report 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., Flanders, the Dutch speaking part of Belgium, and the Netherlands) (e.g., Grondelaers, Van Hout & Speelman, 2011; Speelman, Spruyt, Impe & Geeraerts, 2013; Preston, 2016). 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[1], while 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 situate the study from both perspectives.

1.1. Methodological Perspective

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). After a period of limited innovation, linguistic attitude research now seems to be catching up (Preston, 2016). Examples of studies providing new impulses for traditional language attitude research are Grondelaers & Speelman (2015), who apply new techniques to analyze responses from keyword tasks, Staum Casasanto, Grondelaers & Van Hout (2015), who use images in a forced choice task to replace traditional Likert scale ratings with verbal descriptions of social traits, or Montgomery & Stoeckle (2013), who introduce innovations in processing data collected through 'draw a map' tasks. Not only are linguists starting to improve and refine traditional methods, they are also gaining interest in exploring those measures recently developed in social psychology.

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, McGhee & Schwartz, 1998; see Teige-Mocigemba, Klauer & Sherman, 2010 for a more recent introduction). The IAT is a reaction time based categorization task that measures the association between two binary concepts (e.g., candy/vegetables and good/bad). So far linguists have employed the IAT to study the following aspects of language varieties and linguistic variants: their evaluation (Babel, 2010; Redinger, 2010; Pantos & Perkins, 2012; Lee, 2015; Watt & Llamas, 2015; Loudermilk, 2015; Leinonen, 2016), their social meaning and indexicality (Campbell-Kibler, 2012, 2013; Llamas, Watt & MacFarlane, 2016; Hilton,

Rosseel, Smidt & Coler, 2016), and their salience (Leinonen, 2016).

The IAT is used very frequently in social psychology and has been successfully applied to the study of a wide variety of topics, ranging from racial stereotypes (e.g., Greenwald et al., 1998) to addictive behaviour (e.g., Houben & Wiers, 2006) or advertising (e.g., Maison, Greenwald & Bruin, 2004). One reason for the IAT's popularity in social psychology is that it has been shown to have good psychometric qualities (Nosek, Greenwald & Banaji, 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). An introductory overview of the IAT for a linguistic audience can be found in Rosseel, Geeraerts & Speelman (2014). 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, Holland & Van Knippenberg, 2004) and the Single Attribute IAT (Penke, Eichstaedt & Asendorpf, 2006) allow for non-binary concepts.

Another problem the IAT is claimed to suffer from is the measurement of extra-personal associations instead of personal associations (although this distinction is not uncontroversial. See Gawronski, Peters & LeBel, 2008). Resonating with Karpinski & Hilton's (2001) concept of environmental associations, Olson & Fazio (2004: 653) describe these extra-personal associations as 'associations that are available in memory but are irrelevant to the perceived likelihood of personally experiencing a positive or negative outcome on interaction with the attitude object'. Hence, personal associations refer to preferences endorsed by an individual. Extra-personal associations, on the other hand, are present in memory, because they are frequently encountered in society, 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 vegetables, because this 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 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.

The P-IAT 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:

I like vs. I don't like) by comparing reaction times in a number of computer-based categorization tasks. For each of these four concept categories, a number of stimuli is required that are representative of their respective category.

In a P-IAT, participants are asked to categorize 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 1). A P-IAT is made up of a series of trials which each require the categorization 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 categorization 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 1). The second practice block involves the categorization of attribute stimuli according to the attribute categories 'I like' vs. 'I don't like' (see block 2 in Figure 1). 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 categorized in these blocks using the response keys on which both a target and an attribute category are currently mapped (e.g. 'variety A' + 'I don't like' for the left-hand key, and 'variety B' + 'I like' for the right-hand key, see block 3-4 in Figure 1). Note that each stimulus belongs to one category 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 now swapped sides (see block 5 in Figure 1). 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[2] (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 categorized. 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 block 6 and 7
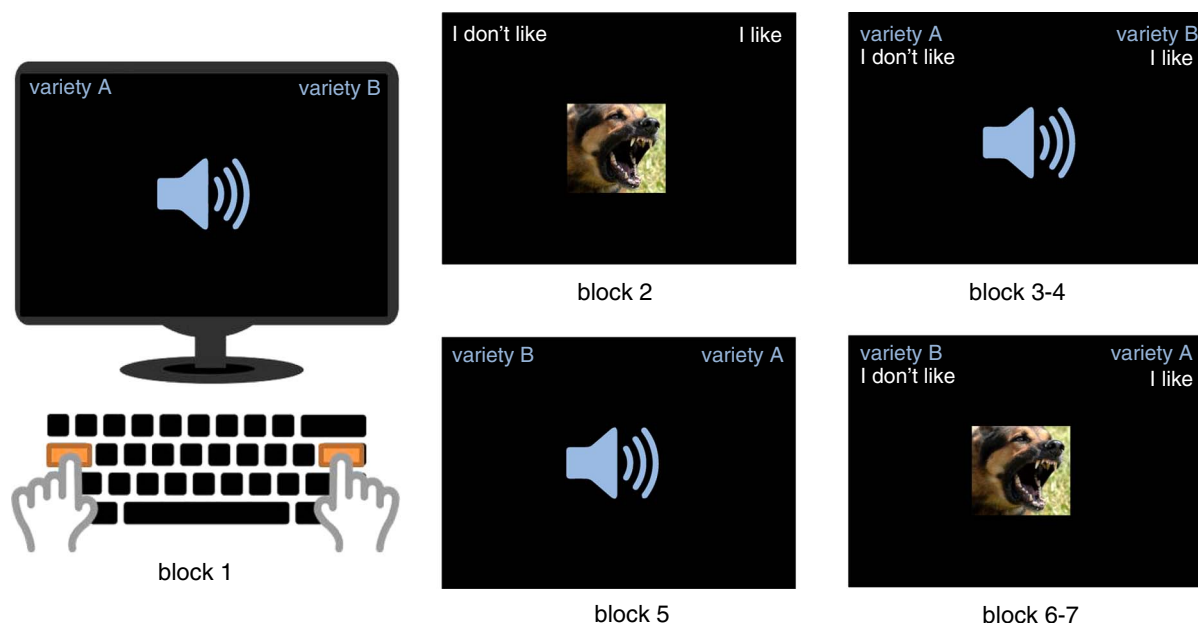
Figure 1. Screenshots with an example of a trial from each block of a P-IAT. The example for block 1 also illustrates the experimental set-up.

that is the reverse of the mapping in block 3 and 4, the other set of experimental blocks (see block 6-7 in Figure 1).

This inverse response key mapping in the two sets of experimental blocks is primordial to the mechanism behind the P-IAT. Categorization 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, categorization 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 stimuli representing variety A and intrinsically positive stimuli, like pictures of a smiling child or a sunny beach, are categorized with the same key. Yet, it will be more difficult for that person if 'variety A' and 'I don't like' are assigned to the same response key. Easier categorization will lead to faster reaction times, while a harder categorization task will slow down responses. By comparing reaction times in the two sets of experimental blocks, we can determine which concepts participants associate more strongly.

In order to personalize the IAT, Olson and Fazio (2004) suggest three changes to the traditional IAT that aim to reduce the activation of extra-personal associations. Firstly, the P-IAT uses attribute labels that refer directly to the participant's opinions, such as 'I like' and 'I don't like' instead of more normative options like 'good' and 'bad' or 'pleasant' and 'unpleasant'. Secondly, corrective error feedback is usually left out in

personalized versions of the IAT, because it may suggest to the participants that they are to categorize items according to societal norms, rather than their own attitudes. A third option for reducing the influence of extra-personal associations is the use of attribute stimuli that are not perceived as universally positive or negative, again to avoid conveying the impression that categorization of the attribute stimuli is to be done based on societal norms rather than personally held attitudes. Instead, attribute items are chosen that are not neutral, but whose valence may differ from participant to participant. It was shown, however, that the former two adaptations suffice to personalize the IAT (Olson & Fazio, 2004: 664).

Except for the studies using the IAT mentioned above, linguists' exploration of other social psychological measures has been rather limited. One such other measure is affective priming (AP, Fazio, Sanbonmatsu, Powell & Kardes, 1986). In an AP experiment, participants are asked to categorize target stimuli as positive or negative. These stimuli are preceded by valenced prime stimuli, often presented subliminally. Depending on whether the valence of the prime is congruent with that of the target, categorization will be faster (in case of congruence: positive prime + positive target, or negative prime + negative target) or slower (in case of incongruence: positive prime + negative target, or negative prime + positive target). Reaction times can then be used to determine the valence of the primes under study. So far, the only research we know of that successfully used AP to measure language attitudes is Speelman et al. (2013).

In this paper, 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 implicit (AP) and one explicit (a rating task using semantic differential scales).

1.2. Descriptive Perspective

Before presenting an overview of the current state of attitudinal studies on varieties of Belgian Dutch and discussing the descriptive aims of this study, let us briefly outline the stratificational structure of Dutch in Belgium. Dutch is the official language in Flanders, the northern part of Belgium. Today, Belgian Dutch is generally taken to represent a situation of diaglossia, to use Auer's (2005) typology (e.g., Geeraerts & Van de Velde, 2013). This means that the linguistic situation can be described as a continuum with on one extreme the base dialects, and on the other, Standard Belgian Dutch (SBD). Although the standard variety of Dutch in Belgium is perceived today as a distinct variety from Standard Netherlandic Dutch, the former was modelled after the latter and spread top-down during a standardization process which started in the nineteenth century, but only gained momentum after the Second World War (Geeraerts, 2017; Marynissen, 2017). The space between the local dialects and SBD in the diaglossic continuum is filled by a highly heterogeneous variety with a regional flavor, often referred to as *tussentaal*. This colloquial variety, which we will refer to in this paper as Colloquial Belgian Dutch (CBD), has proven hard to define and delineate and is to be situated somewhere between the local dialects and Standard Belgian Dutch. This explains why the variety is commonly 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).

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 pick up 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 CBD and its relation to the standard variety and local dialects (Grondelaers, 2013). Overall, this research is rather fragmented, focusing on various regional varieties and participant groups. A large-scale survey of the attitudinal landscape in Dutch speaking Belgium is lacking as yet.

The few recent attitudinal studies carried out in Dutch speaking Belgium, all dealing with CBD compared to SBD[3], 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 plays a mediating role in this trend: the more features, the less status (Impe & Speelman, 2007). On the other hand, CBD is perceived more positively on the solidarity 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 did 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 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 solidarity dimension. It is important to keep in mind though, that the methods, designs, and varieties of CBD investigated 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, Speelman and Geeraerts (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 also is 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 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 our study does not allow us to distinguish between different dimensions of language attitudes, 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 1.1 and 1.2, we can now specify the aims of the present study with regard to both dimensions. From a methodological perspective, 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, Deutsch & Banse, 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, especially when it comes to reliability (Spruyt, Gast & Moors, 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 attitudinal patterns observed largely coincide between the three measures, but are not identical. In the discussion section, we will consider a number of potential explanations of why there is no perfect overlap.

From a descriptive point of view, this paper aims to contribute to a picture of the attitudinal landscape of Belgian Dutch, 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 area, 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 presented itself as a good starting point for exploring the P-IAT as a language attitudes 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). 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,

Table 1. *Hypotheses*

	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)

based on what was reported by Speelman and collea-
gues (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 summarized in Table 1.

2. METHOD

The study consisted of two parts: a P-IAT and a ques-
tionnaire. After giving informed consent, participants
were asked to take part in a P-IAT experiment mea-
suring implicit attitudes. The indirect attitude mea-
surement was followed by a direct one using semantic
differential scales and a short questionnaire collecting
basic demographic information. Afterwards, partici-
pants 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 uni-
versity 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
on 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 lin-
guistics or psychology students were allowed to take the
experiment to avoid participants with previous experi-
ence with either the method or the topic of the study.

Data of 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 they 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[4]. 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 com-
paring reaction times in a series of categorization tasks.
The test used in this study was designed as described in
section 1.1 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 2. The aspects personalizing the P-IAT in
this study were, on the one hand, the use of attribute
labels referring to the participant's subjective opinion,
and on the other, the omission of error feedback for
the attribute stimuli. With these adaptations, we aimed
to make sure participants categorized the attribute
stimuli based on their personal opinions rather than
according to a normative distinction between good and
bad. Note that error feedback was retained for the target
stimuli. This was done for two reasons. Firstly, we
wanted participants to pay attention to the sound
samples and the varieties they represented rather than
guess. Secondly, participants may have experienced the
task of classifying the language varieties under study
based on their own opinion as artificial and counter-
intuitive. Most non-linguists in Belgium conceptualize
the standard variety of Dutch as a neatly delineated,
codified and homogenous variety (e.g. Lybaert, 2014).
Hence, respondents may have reacted surprised or con-
fused if they had been told to categorize the regional and
standard variety in the experiment according to their
personal inclination, given that in their conceptualization

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

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[5]	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. *Stimulus set*

	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 4. *Between subject design of the implicit attitude measurement including participant numbers*

		A *Antwerp* *SBD* *order 1*	B *Antwerp* *SBD* *order 2*	C *West-Flemish* *SBD* *order 1*	D *West-Flemish* *SBD* *order 2*
	target varieties mapping response keys				
origin participants	*Antwerp* *West-Flanders*	23 24	22 21	22 20	23 21

there is no discussion about what constitutes the standard variety, rendering personal opinions irrelevant.

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 (Greenwald, Nosek & Banaji, 2003). We calculated D scores using the IAT package in R (Martin, 2014) which is based on the algorithm described in Greenwald et al. (2003). D scores were analysed using multiple linear regression. This method was deemed most robust given the slightly unbalanced sample (see Table 4).

The stimuli and category labels used in the study are summarized in Table 3. 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 live in the region of which they produced the accent. They both have a clear and pleasant voice, which is neither nasal, creaky, whispery nor harsh (Laver, 1994; Impe, 2010). No differences in speech rate were found between the speakers (Street, Brady & Lee, 1984; Impe, 2010) and the recordings were made in a professional radio studio in order to ensure good sound quality. 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, Schouppe, Hoekstra & Van de Wouden, 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). To ensure the target stimuli were representative for their respective varieties, a small panel of native speakers was consulted (see Impe, 2010). Furthermore, no participants had to be excluded from the study based on high error rates or slow responses in the categorization of the target stimuli (see Greenwald et al., 2003 for exclusion criteria based on error rates and fast/slow latencies). This suggests that the stimuli were readily identifiable as belonging to the three varieties under study. No participants reported any problems regarding representativeness of the audio stimuli when given the opportunity to comment on any aspect of the experiment at the end of the study. Note that the target stimuli only differ in the accent with which they are produced. Only standard lexical items commonly found in both SBD and CBD were used. No regional or dialectal words figured in the stimulus set.

For the attribute stimuli, we used 5 positive and 5 negative real life colour pictures for which norm data have been collected (Spruyt, Hermans, De Houwer & Eelen, 2002). All pictures were equal in size (410x308 pixels). The experiment was run on a laptop

with a screen resolution of 1366x768 using Affect 4.0 (Spruyt, Clarysse, Vansteenwegen, Baeyens & Hermans, 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/slecht* (literally 'I find it good/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.[6]

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, participants 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). 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.

In the between subject design, summarized 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 as 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.

Note that we chose to start the study with the P-IAT rather than with the explicit rating task. This was done in order to avoid the possibility that the rating task would activate certain (normative) associations before participants started the P-IAT, as well as to conceal the aim of the study as much as possible. Respondents were told that they were participating in an experiment about the processing of images and sound in the brain. However, given the fact that the target and attribute categories are clearly communicated during the P-IAT and the direct nature of the explicit rating task that followed afterwards, no doubt many participants will have had some idea of the aim of the study. Nonetheless, multiple participants expressed surprise when given more information about the goal of the study at the end of the experiment. A potential disadvantage of presenting the P-IAT and rating task in this order is a diminished correlation between the results of both experimental tasks. We will come back to this issue in Section 4.2 below.

3. Results

3.1. Implicit Attitude Measurement

After computing D scores based on the reaction times measured in the experimental blocks, we entered these difference scores into a linear regression analysis as the response variable and participants' region of origin and language varieties included in the P-IAT as the predictor variables. The resulting model is summarized in Table 5. Note that sum coding was used, so the estimate for the intercept represents the grand mean.

The grand mean reported in Table 5 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 2, which summarizes 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 higher than 0. For Antwerp participants in experiment

AB: $M = 0.24$, $SE = 0.04$, $t(44) = 5.88$, $p < 0.0001$; for West-Flemish participants in experiment AB: $M = 0.36$, $SE = 0.05$, $t(44) = 7.43$, $p < 0.0001$; for Antwerp participants in experiment CD: $M = 0.39$, $SE = 0.04$, $t(44) = 10.30$, $p < 0.0001$; for West-Flemish participants in experiment CD: $M = 0.14$, $SE = 0.05$, $t(40) = 2.92$, $p < 0.0001$).

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 5). If we tease apart the interaction using post hoc tests (*t*-tests with Bonferroni correction), we find that Antwerp participants show a stronger preference ($M = 0.39$, $SE = 0.04$) compared to their West-Flemish counterparts ($M = 0.14$, $SE = 0.05$) in the CD experiment which contained the standard

Table 5. *Summary linear regression model of D scores using sum coding*

predictor	estimate	p	
intercept (grand mean)	0.281	<0.0001	***
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	<0.0001	***

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

variety and West-Flemish regiolect, $t(78.83) = 4.21$, $p < 0.0001$. 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 participants' 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: $M = 0.39$, $SE = 0.04$; $t(87.52) = −2.70$, $p < 0.05$) and West-Flemings (experiment AB: $M = 0.36$, $SE = 0.05$; experiment CD: $M = 0.14$, $SE = 0.05$, $t(40) = 2.92$; $t(83.94) = −3.27$, $p < 0.01$). This pattern can be described as a decrease of participants' preference for SBD when presented with their own regiolect and hence can 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 contextualized by the specific comparisons of target concepts included in the experiment.

To summarize, 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.
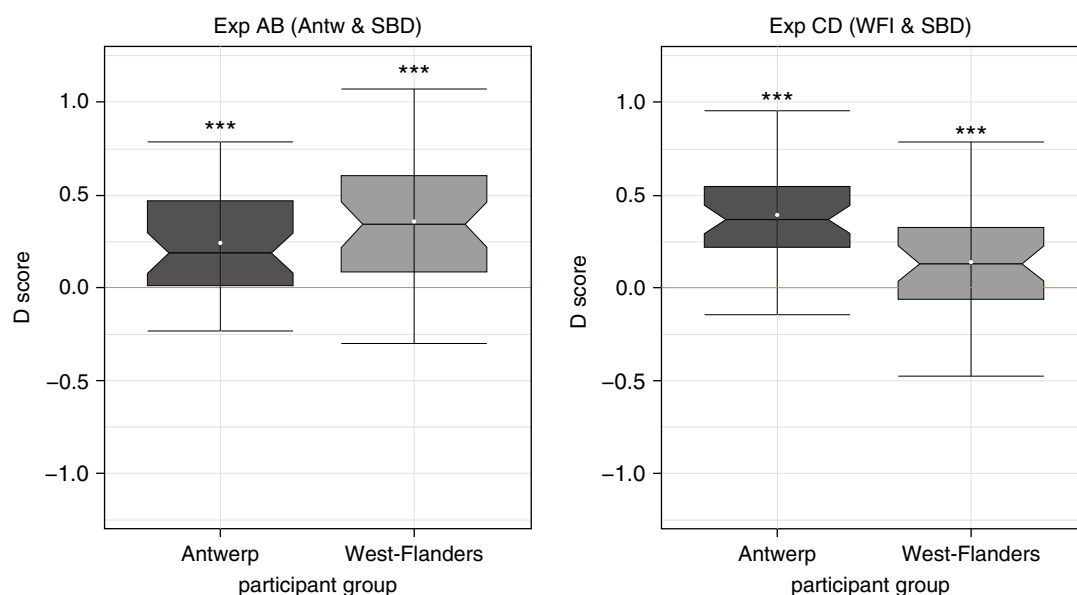

Figure 2. Boxplots of D-scores in experiment AB and experiment CD. Positive D-scores indicate a preference for the standard, negative D-scores a preference for the regional variety included in the experiment.
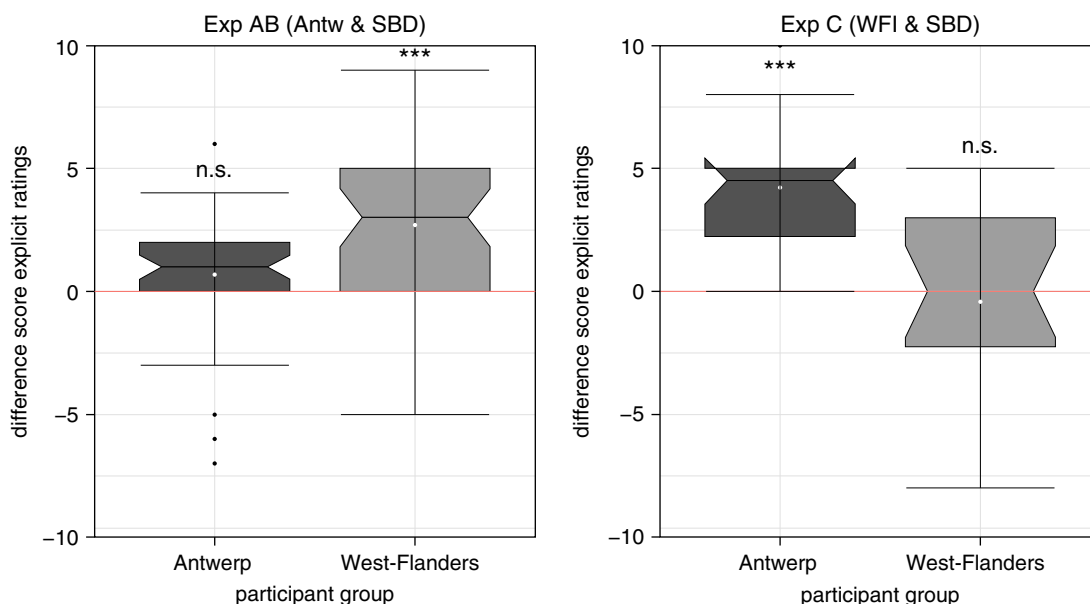
Figure 3. Boxplots displaying the difference scores between the explicit ratings for SBD and the regional variety in experiments AB and C. Positive scores indicate a preference for the standard, negative scores a preference for the regional variety included in the experiment.

3.2. Explicit Attitude Measurement

The results of the explicit attitude measurement are summarized in Figure 3 and show a similar pattern to the one in the implicit measurement. Note that due to a technical problem with the display of the question accompanying the rating scale, the explicit attitudes of participants taking the D experiment 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 rating of SBD and the regional variety presented in the experiment, in order to make both measures as comparable as possible.

As for the implicit measurement, the outcome of the explicit measurement was modeled using linear regression (with sum coding). The summary of the linear regression model (see Table 6) shows that the grand mean is significantly larger than 0 which is an indication of a general preference for the standard variety. 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 ($M = 2.71$, $SE = 0.46$, $t(44) = 286$, $p < 0.0001$) and Antwerp participants in experiment C ($M = 4.06$, $SE = 0.62$, $t(17) = 286$, $p < 0.0001$). 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

Table 6. *Summary linear regression model of explicit attitude ratings using sum coding*

predictor	estimate	p	
intercept (grand mean)	1.83	<0.0001	***
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	<0.0001	***

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

presented with the other group's regiolect, just as we observed for the P-IAT results.

In the regression model presented in Table 6[7], the interaction between variety and participant origin is highly significant. 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: $M = 2.71$, $SE = 0.46$, $t(85.51) = −3.20$, $p < 0.01$) and experiment C (Antwerpians: $M = 4.23$, $SE = 0.51$; West-Flemings: $M = -0.4$, $SE = 0.84$, $t(31.74) = 4.68$, $p < 0.001$) 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

Table 7. *Correlations between implicit and explicit attitude measures*

experiment	participant group	r_s	p	
AB	Antwerp	0.275	0.067	.
	West-Flanders	−0.017	n.s.	
C	Antwerp	−0.091	n.s.	
	West-Flanders	0.424	0.062	.

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

with the other group's variety. (Antwerp participants: $M_{AB} = 0.78$, $SE_{AB} = 0.39$; $M_C = 4.23$, $SE_C = 0.51$, $t(44.96) = -5.35$, $p < 0.0001$; West-Flemish participants: $M_{AB} = 2.71$, $SE_{AB} = 0.46$; $M_C = -0.4$, $SE_C = 0.84$, $t(30.91) = 3.23$, $p = 0.01$). This 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.

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 7). In both cases the correlation just misses 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.

4. DISCUSSION

In this section, we will have a closer look at the results (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 (4.3). The discussion of the results will be approached in two steps. First, in section 4.1, we will 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 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 of 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 the two studies 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: auditory fragments representing actual language use), the IAT measures attitudes towards both the stimuli in the experiment (in our study: the same auditory fragments) *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, Coupland & Garrett, 2005; Kristiansen, 2010; Grondelaers & Kristiansen, 2013). Coupland & Bishop (2007:85) suggest that attitudes towards labeled 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 perspective more salient. Considering this explicit comparison of varieties against the background of the normative nature of SBD that resulted from the top-down standardization 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 and 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 is out of our control.

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 (e.g. Hofmann, Gawronski, Gschwendner, Le & Schmitt, 2005a; Nosek, Greenwald & Banaji, 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, Swann & Pennebaker (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 they 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, Gschwendner, Nosek & Schmitt, 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, Heider, Spruyt, Roets & Hughes, 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, Burkley & Stokes, 2008). This term refers to whether or not two methods measure the same type of construct. As pointed out in 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 modest or lacking correlations between the implicit and explicit measurements may reside in the degree of social sensitivity of the domain under study. Greenwald, Poehlman, Uhlmann & Banaji (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[8]. Besides, social sensitivity 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 variety, but no correlation whatsoever when measuring their attitudes towards another group's 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.

To conclude this discussion of the correlation between our implicit and explicit attitude measurements, we would like to emphasize that one needs to be careful drawing conclusions about the nature of implicit vs. explicit attitudes based on (lacking) correlations between measures (cf. Gawronski & De Houwer, 2014). Although a (lack of) correlation between measures could have theoretical significance, the discussion above clearly shows that methodological explanations cannot be excluded.

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? This evaluation of the P-IAT will be embedded in a discussion of the IAT paradigm at large.

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, Greenwald, Brown, Gray & Snowden, 2010), which makes it an interesting option to study associations participants are unwilling to share explicitly or 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 earlier in the paper, 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 of the P-IAT, practical restrictions on large-scale P-IAT experiments, the selection of suitable stimuli, the need for further research on the categorization mechanisms at play in the IAT paradigm, the relevance of (extra)personal associations for language attitude research, 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 attitudes 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, these are incompatible with the use of auditory target stimuli without running into the problem of recoding based on the modality of the stimuli[9] (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 categorized 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 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 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. The advantage of the P-IAT's comparative structure then is that it forces the researcher to make explicit this 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 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 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, Bluemke & Wänke, 2007; Xu, Nosek & Greenwald, 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 of conducting such larger-scale studies (Nosek, Banaji & Greenwald, 2002a, 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 aspect 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 or familiarity (see for instance Moors, De Houwer, Hermans 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 (for a

discussion of implicitness defined in terms of automaticity, see De Houwer, Teige-Mocigemba & Spruyt, 2009). 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. When working with the IAT paradigm, it is also vital to keep in mind that the linguistic phenomena under study have to be represented by small set of short stimuli. This means the stimuli have to be selected and pretested very carefully if one wants to be certain they are representative for the phenomenon under study, especially if the stimuli represent an entire variety as was the case in the study reported here.

The length restriction also entails a theoretical-issue: the language stimuli in a P-IAT are completely decontextualized. Hence, one could question the validity of a language attitudes measure if the language presented in the experiment is decontextualized to such a high degree. However, if a memory component of attitudes is assumed—as it is by many psychologists (Albarracín, Wang & Noguchi, 2008) as well as linguists (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, 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 decontextualization. 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 limited research available in social psychology on this topic seems encouraging (e.g., Gschwedner, Hofmann & Schmitt, 2008 for racial attitudes and the IAT as an anxiety measure), and we are currently conducting experiments to explore the possibility of including situational context in the P-IAT as a language attitudes 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 categorization 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.

As noted above, the discussion of the P-IAT in this section is part of a discussion of the IAT paradigm at large. This is justified given that most of the structural and procedural aspects discussed in this section are highly similar for the P-IAT and the traditional IAT. For a discussion of the aspects that set the former apart from the latter, we refer to section 1.1 above where it was explained that the P-IAT aims to avoid measuring extrapersonal associations. However, given that this is the first study to apply the P-IAT to measure language attitudes, not much is known about the role of extrapersonal vs. personal associations in the domain of language attitudes. Hence, we propose that future research further investigates the role of these concepts in the light of sociolinguistic attitude research, which could be accomplished by comparing results from P-IAT and traditional IAT experiments. Such studies would both enhance our methodological understanding of the linguistic P-IAT compared to the linguistic non-personalized IAT, as well as our theoretical understanding of the concept of (extra-)personal associations and their role in the perception of language variation.

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 paper 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 very 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 the 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 comparing results from different measures.

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 disposal 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) 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.

Acknowledgments

Notes

[1] In this paper, we opted to use the term 'attitudes', because this term is commonly used in both linguistics and social psychology. However, we would like to stress that we do not interpret the term 'language attitudes' in a strict way as referring to evaluation only. By contrast, we use it in the broadest sense as the social meaning of a language variety or variant (cfr. Soukup, 2013).

[2] The IAT has been shown to suffer from block order effects: if the congruent block precedes the incongruent block, 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, Klauer & Rothermund, 2008; the Recoding Free IAT, IAT-RF, Rothermund, Teige-Mocigemba, Gast & Wentura, 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 this study, 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.

[3] Except for Grondelaers et al. (2011) who did not include SBD in their study and focused on a variety in between fully-fledged 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.

[4] All analyses were conducted with and without these two outliers and led to the same conclusions, the only difference being that the Bonferroni corrected post hoc *t*-test comparing participants from Antwerp and West-Flanders in the AB P-IAT experiment just misses significance when the outliers are removed from the data set (see section 3.1), while it reaches significance if these outliers are retained. Hence, it was decided to report the analyses without these two outliers.

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

[6] As a reviewer rightly pointed out, 'neutral accent' is not a value-free label which may or may not have influenced participants' reactions in the P-IAT in the direction of a pro-standard bias. However, we chose to use this label given that other alternatives such as 'standard accent', 'AN' (*Algemeen Nederlands*, 'General Dutch') or 'ABN' (*Algemeen Beschaafd Nederlands*, 'Generel Educated Dutch') are heavily ideologically charged in the Belgian context. Hence we suspect these alternatives would have led to a much stronger pro-standard bias due to their association with normativity. This issue clearly demonstrates how the necessity to provide labels in the P-IAT may prove problematic in certain cases. Future studies should investigate the influence of the use of different labels on the associations measured in the P-IAT. See also section 4.3 for a further discussion on the use of labels in the P-IAT.

[7] Given that 'origin participant' interacts significantly with 'variety', which makes the main effect of the former less informative (Baayen 2008, 166), 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.

[8] 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) 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 ($r = .17$, $p = .086$; $r = .43$, $p = .077$).

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

References

Albarracín, Dolores, Wei Wang, Hong Li & Kenji Noguchi. 2008. Structure of attitudes. Judgments, memory, and implications for change. In William D. Crano & Radmila Prislin (eds.), *Attitudes and Attitude Change*, 19-39. New York: Psychology Press.

Baayen, R. Harald. 2008. *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: CUP.

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

Bezooijen, Renée van. 2001. Poldernederlands. Hoe kijken vrouwen ertegenaan? *Nederlandse Taalkunde*, 6. 257-271.

Bishop, Hywel, Nikolas Coupland & Peter Garrett. 2005. Conceptual accent evaluation: Thirty years of accent prejudice in the UK. *Acta Linguistica Hafniensia: International Journal of Linguistics*, 37. 131-154.

Bosson, Jennifer K., William B. Swann & James W. Pennebaker. 2000. Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology*, 79(4). 631-643.

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

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

Campbell-Kibler, Katherine. 2013. Connecting attitudes and language behavior via implicit sociolinguistic cognition. In Tore Kristiansen & Stefan Grondelaers (eds.), 307–329.

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

Cuvelier, Pol. 2007. Standaardnederlands, tussentaal en dialect in Antwerpen. De perceptie van jonge moedertaalsprekers en taalleerders. In Dominiek Sandra, Rita Rymenans, Pol Cuvelier & Peter Van Petegem (eds.), *Tussen taal, spelling en onderwijs. Essays bij het emeritaat van Frans Daems*, 39-58. Gent: Academia Press.

Cvencek, Dario, Anthony G. Greenwald, Anthony S. Brown, Nicola S. Gray & Robert J. Snowden. 2010. Faking of the implicit association test is statistically detectable and partly correctable. *Basic and Applied Social Psychology*, 32. 302-314.

De Houwer, Jan, Niclas Heider, Adriaan Spruyt, Arne Roets & Sean Hughes. 2015. The relational responding task: Toward a new implicit measure of beliefs. *Frontiers in Psychology* 6. Article, 319.

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

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

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

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

Garrett, Peter. 2005. Attitude measurements. In Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier, & Peter Trudgill (eds.), *Sociolinguistics: An international handbook of the science of language and society*, vol. 3.2, 1251-1260. Berlin: Mouton De Gruyter.

Gast, Anne & Klaus Rothermund. 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, Bertram, Roland Deutsch & Rainer Banse. 2011. Response Interference Tasks as Indirect Measures of Automatic Associations. In Karl C. Klauer, Andreas Voss & Christoph Stahl (eds.), *Cognitive Methods in Social Psychology*, 78-123. New York: Guilford Press.

Gawronski, Bertram, Kurt R. Peters & Etienne P. LeBel. 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, Bertram & Jan De Houwer. 2014. Implicit measures in social and personality psychology. In Harry T. Reis & Charles M. Judd (eds.), *Handbook of Research Methods in Social and Personality Psychology*, 2nd ed., 283-310. New York: Cambridge University Press.

Geeraerts, Dirk. 2017. Het kegelspel der taal. De naoorlogse evolutie van de Standaardnederlandsen. In Gert De Sutter (ed.), 100-120.

Geeraerts, Dirk & Hans Van de Velde. 2013. Supra-regional characteristics of colloquial Dutch. In Frans Hinskens & Johan Taeldeman (eds.), 532-556.

Ghyselen, Anne-Sophie. 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.

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

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

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

Greenwald, Anthony G., T. Andrew Poehlman, Eric Luis Uhlmann & Mahzarin R. Banaji. 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, Stefan. 2013. Attitude measurements in the Low Countries. In Frans Hinskens & Johan Taeldeman (eds.), 586–602.

Grondelaers, Stefan & Tore Kristiansen. 2013. On the need to access deep evaluations when searching for the motor of standard language change. In Tore Kristiansen & Stefan Grondelaers (eds.), 9–52.

Grondelaers, Stefan & Dirk Speelman. 2013. Can speaker evaluation return private attitudes towards stigmatised varieties? Evidence from emergent standardisation in Belgian Dutch. In Tore Kristiansen & Stefan Grondelaers (eds.), 171–191.

Grondelaers, Stefan & Dirk Speelman. 2015. A quantitative analysis of qualitative free response data. Paradox or new paradigm? In Jocelyne Daems, Eline Zenner, Kris Heylen, Dirk, Speelman & Hubert Cuykens (eds.), *Change of paradigms – new paradoxes: Recontextualizing language and linguistics*, 361-384. Berlin/Boston: De Gruyter Mouton.

Grondelaers, Stefan & Roeland van Hout. 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, Stefan, Roeland van Hout & Dirk Speelman. 2011. A perceptual typology of standard language situations in the Low Countries. In Nikolas Coupland & Tore Kristiansen (eds.), *Standard languages and language standards in a changing Europe,* 199–222. Oslo: Novus.

Gschwendner, Tobias, Wilhelm Hofmann & Manfred Schmitt. 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.

Hilton, Nanna Haug, Laura Rosseel, Eva M. Smidt & Matt Coler. 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.

Hofmann, Wilhelm, Bertram Gawronski, Tobias Gschwendner, Huy Le & Manfred Schmitt. 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, Wilhelm, Tobias Gschwendner, Brian A. Nosek & Manfred Schmitt. 2005b. What moderates implicit—explicit consistency? *European Review of Social Psychology*, 16(1). 335-390.

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

Impe, Leen. 2006. *Een attitudineel mixed guise-onderzoek naar tussentaal in Vlaanderen*. Leuven: University of Leuven MA dissertation.

Impe, Leen. 2010. *Mutual intelligibility of national and regional varieties of Dutch in the Low Countries*. Leuven: University of Leuven dissertation.

Impe, Leen & Dirk Speelman. 2007. Vlamingen en hun (tussen)taal: Een attitudineel mixed guise-onderzoek. *Handelingen van de Koninklijke Zuid-Nederlandse Maatschappij Voor Taal- En Letterkunde*, 16. 109-128.

Karpinski, Adrew & James L. Hilton. 2001. Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology*, 81. 774-778.

Karpinski, Andrew & Ross B. Steinman. 2006. The Single Category Implicit Association Test as a measure of implicit

social cognition. *Journal of Personality and Social Psychology*, 91(1). 16-32.

Kristiansen, Tore. 2010. 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.

Laver, John. 1994. *Principles of Phonetics*. University Press Cambridge.

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

Leinonen, Therese. 2016. Attitudes, salience, and accommodation – quantity in Finland-Swedish. Talk presented at *ExAPP 2016*, 22 September 2016, Vienna.

Llamas, Carmen, Dominic Watt & Andrew E. MacFarlane. 2016. Estimating the relative sociolinguistic salience of segmental variables in a dialect boundary zone. *Frontiers in Psychology*, 7. 1163.

Loudermilk, Brandon C. 2015. Implicit attitudes and the perception of sociolinguistic variation. In Alexei Prikhodkine & Dennis Preston (eds.), 137-156.

Lybaert, Chloé. 2014. *Het gesproken Nederlands in Vlaanderen. Percepties en attitudes van een spraakmakende generatie*. Gent: Universiteit Gent dissertation.

Maison, Dominika, Anthony G. Greenwald & Ralph H. Bruin. 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, Dan. 2014. *IAT: Functions to use with data from the Implicit Association Test*. R package version 0.2. http://CRAN.R-project.org/package=IAT.

Marynissen, Ann. 2017. De lange weg naar een Nederlandse standaardtaal. Een beknopte geschiedenis van de standaardisering van het Nederlands. In Gert De Sutter (ed.), 60-79.

McKenzie, Robert. 2017. Implicit associations and explicit attitudes towards Northern English and Southern English speech. Invited talk at *Newcastle University Applied Linguistics Seminar Series*, 7 March 2017, Newcastle, UK.

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

Moors, Agnes, Jan De Houwer, Dirk Hermans, Sabine Wanmaker, Kevin van Schie, Anne-Laura Van Harmelen, Maarten De Schryver, Jeffrey De Winne & Marc Brysbaert. 2013. Norms of valence, arousal, dominance, and age of acquisition for 4300 Dutch words. *Behavior Research Methods*, 45(1). 169-177.

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

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

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

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

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

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

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

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

Preston, Dennis R. 2015. Does language regard vary? In Alexei Prikhodkine & Dennis R. Preston (eds.), 3-36.

Preston, Dennis R. 2016. Whaddayaknow now? In Anna Babel (ed.), *Awareness and control in sociolinguistic research*, 177-199. Cambridge: CUP.

Redinger, Daniel. 2010. *Language attitudes and code-switching behaviour in a multilingual educational context : The case of Luxembourg*. York: University of York dissertation.

Rosseel, Laura, Dirk Geeraerts & Dirk Speelman. 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.

Rothermund, Klaus, Sarah Teige-Mocigemba, Anne Gast & Dirk Wentura. 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.

Schuurman, Inneke, Machteld Schouppe, Heleen Hoekstra & Ton van der Wouden. 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*, 101-108. Budapest.

Soukup, Barbara. 2013. The measurement of "language attitudes" - A reappraisal from a constructionist perspective. In Tore Kristiansen & Stefan Grondelaers (eds.), 251–266.

Speelman, Dirk, Adriaan Spruyt, Leen Impe & Dirk Geeraerts. 2013. Language attitudes revisited: Auditory affective priming. *Journal of Pragmatics*, 52. 83-92.

Spruyt, Adriaan, Jeroen Clarysse, Debora Vansteenwegen, Frank Baeyens & Dirk Hermans. 2010. Affect 4.0: a free software package for implementing psychological and psychophysiological experiments. *Experimental Psychology*, 57(1). 36-45.

Spruyt, Adriaan, Anne Gast & Agnes Moors. 2011. The sequential priming paradigm: A primer. In Karl C. Klauer et al. (eds.), 48–77.

Spruyt, Adriaan, Dirk Hermans, Jan De Houwer & Paul Eelen. 2002. On the nature of the affective priming effect: Affective priming of naming responses. *Social Cognition*, 20(3). 227-256.

Staum Cassasanto, Laura, Stefan Grondelaers & Roeland van Hout. 2015. Got Class? Community-shared conceptualizations of social class in evaluative reactions to sociolinguistic variables. In Alexei Prikhodkine & Dennis R. Preston (eds.), 159-173.

Steffens, Melanie C. 2004. Is the implicit association test immune to faking? *Experimental Psychology*, 51(3). 165-179.

Street, Richard L. Jr., Robert M. Brady & Raymond Lee. 1984. Evaluative responses to communicators: The effects of speech rate, sex and interaction context. *The Western Journal of Speech Communication*, 48(1). 14-27.

Teige-Mocigemba, Sarah, Karls C. Klauer & Klaus Rothermund. 2008. Minimizing method-specific variance in the IAT: A Single Block IAT. *European Journal of Psychological Assessment*, 24(4). 237-245.

Teige-Mocigemba, Sarah, Karl C. Klauer & Jeffrey W. Sherman. 2010. A practical guide to Implicit Association Test and related tests. In Bertram Gawronski & Keith Payne (eds.), *Handbook of implicit social cognition: Measurement, theory and applications*, 117-139. New York: Guilford.

Vandekerckhove, Reinhild & Pol Cuvelier. 2007. The perception of exclusion and proximity through the use of standard Dutch, "tussentaal" and dialect in Flanders. In Pol Cuvelier, Theodorus du Plessis, Michael Meeuwis & Lut Teck (eds.), *Mulitlingualism and exclusion. Policy, practice and prospects*, 241-256. Hatfield, Pretoria: Van Schaik.

Van Gijsel, Sofie, Dirk Speelman & Dirk Geeraerts. 2008. Style shifting in commercials. *Journal of Pragmatics*, 40(2). 205-226.

Watt, Dominic & Carmen Llamas. 2015. Perception of difference: Socioindexical forms in the Scottish/English border region. Talk presented at *ICLaVE 8*, 27 May 2015, Leipzig.

Wigboldus, Daniel H. J., Robert W. Holland & Ad van Knippenberg. 2004. *Single target implicit associations*. Unpublished manuscript.

Xu, Kaiyuan, Brian Nosek & Anthony G. Greenwald. 2014. Data from the Race Implicit Association Test on the Project Implicit Demo Website. *Journal of Open Psychology Data*, 2(1). e3.