The relational responding task (RRT): a novel approach to measuring social meaning of language variation
Rosseel, Laura; Speelman, Dirk; Geeraerts, Dirk

Published in:
Linguistics Vanguard

DOI:
https://doi.org/10.1515/lingvan-2018-0012

Publication date:
2019

Citation for published version (APA):
The relational responding task (RRT): a novel approach to measuring social meaning of language variation

Laura Rosseel
KU Leuven
laura.rosseel@kuleuven.be

Dirk Speelman
KU Leuven
dirk.speelman@kuleuven.be

Dirk Geeraerts
KU Leuven
dirk.geeraerts@kuleuven.be

More than half a century after the introduction of the Matched Guise Technique (MGT, Lambert et al. 1960), sociolinguistic attitude research is witnessing a new influx of attitudes measures developed in social psychology. The Implicit Association Test (IAT, Greenwald et al. 1998) and its variants, for instance, are finding increasing acceptance among linguists studying the social meaning of language variation (e.g. Pantos and Perkins 2012; Campbell-Kibler 2012; McKenzie 2018; Rosseel et al. 2018). Another one of these psychological measures that has been applied to the study of language attitudes is affective priming (Speelman et al. 2013). A feature shared by these new techniques is that they aim to capture implicit processes, i.e. automatic associations between an attitude object and its evaluation.\(^1\) Automaticity refers to processes that occur under circumstances where participants have no control over the process, little time, few cognitive resources, no particular intentions and limited awareness (De Houwer et al. 2009). These implicit processes do not entirely coincide with what linguists tend to refer to as ‘covert’, ‘deep’ or ‘private’ language attitudes.

Although the latter are usually not defined precisely, they mostly refer to attitudes about language that speakers are not aware of (Kristiansen 2010), while awareness is only one of several aspects that can play a role in implicit processes. In this paper, we introduce a new implicit attitude measure that was recently developed in social psychology to the linguistic study of attitudes: the Relational Responding Task (RRT, De Houwer et al. 2015). The RRT is already beginning to be applied successfully in various domains of psychology. Examples are Koning and colleagues (2016) who used the measure to study parenting cognition and alcohol use or Dewitte and collaborators (2017) who study sexual self-concept with the RRT. As a first linguistic exploration of this new method, we apply it to measure associations with two varieties of Belgian Dutch: Standard Belgian Dutch and Colloquial Belgian Dutch, which are hypothesized to be associated with prestige and dynamism respectively.

In what follows, we first introduce the RRT (Section 1) and provide some background information on the language varieties included in the case study (Section 2). Section 3 focuses on the materials and procedure of the case study. The results, then, are presented and discussed in Sections 4 and 5 respectively.

\(^1\) Our understanding of ‘evaluation’ is not restricted to positive and negative valence, as is sometimes the case in social psychology. A fortiori, we define language attitudes in the broadest sense possible and equate it with the social meaning of language (variation) (cf. Soukup 2013). Hence, our use of the term ‘language attitudes’ is similar to Preston’s (2010) umbrella concept of ‘language regard’ which groups notions like ‘attitudes’, ‘beliefs’ and ‘ideologies’.
1. The Relational Responding Task

The RRT is a method that was only introduced very recently (De Houwer et al. 2015), but it is already starting to find wide acceptance among psychologists (e.g. Koning et al. 2016; Dewitte et al. 2017; Walker 2017; Heider et al. 2018). The aim of the RRT is to measure propositional relations, rather than associative ones as is for instance the case in the IAT (De Houwer et al. 2015). Traditionally, many psychologists adhere to a view of memory as consisting of an associative component and a propositional one, with the former being prone to automatic processes, but not the latter, which is considered to operate solely under controlled circumstances (Hughes et al. 2016: 634). De Houwer et al. (2015), however, argue that propositional relations may be activated automatically as well, and developed the RRT to measure these relations. Psychologists refer to these propositional relations as beliefs (De Houwer et al. 2015). To make this more concrete, consider the example where a researcher is interested in studying body image. The IAT could be used in this case to measure the association between the concepts ‘self’ (I vs. other) and ‘weight’ (fat vs. thin). If a participant’s results show an association between the self and thin, the researcher lacks information on how the two concepts are related for this participant. That is to say, in research on body image, the difference between ‘I am thin’ and ‘I want to be thin’ is crucial and the IAT cannot distinguish between the two (see also Pantos, this issue). This is where the RRT can make a difference. The RRT does not measure the association between two loose concepts (like ‘I’ and ‘thin’), rather, it focuses on propositions (like ‘I am thin’ or ‘I want to be thin’). The idea is that participants categorise a number of statements as being true or false, not based on what they think themselves, but reacting as if they adhered to a certain belief (which is imposed on them by the researcher). If this imposed belief matches their own belief, they will be able to categorise the statements faster than when this is not the case. As a result, comparing participants’ reaction times in categorisation tasks with different imposed beliefs can tell us which beliefs they adhere to.

The RRT contains seven blocks of trials: two practice blocks followed by two experimental blocks, another practice block and finally, two experimental blocks again (see Figure 1). The first practice block presents the participants with a series of inducer trials. The stimuli in these trials are usually synonyms of true and false (e.g. correct, right, incorrect, wrong) and have to be categorised as such using two response keys. The meaning of the response keys (‘true’ and ‘false’) is indicated in the top corners of the screen and remains the same throughout the RRT. The main function of the inducer trials is to remind participants of the meaning of the response keys and to avoid response recoding, i.e. categorising based on another feature of the stimuli than whether they are synonyms of true or false (De Houwer et al. 2015; Dewitte et al. 2017).
Figure 1 - Screenshots with an example of a trial from each block of a RRT. The example for block 1 also illustrates the experimental setup.

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

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

Block 5 constitutes another practice block containing the same statements as in blocks 2-4, but no inducer trials. Now participants are instructed to respond to the propositions as if they endorsed rule 2 which states the opposite of rule 1. In our example about language varieties A and B, rule 2 would be ‘variety B sounds more prestigious than variety A’. So now the correct response to ‘variety A sounds more intelligent than variety B’ would be false, and to ‘variety B sounds more educated than variety A’ it would be true.

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

Social psychologists also have another implicit measure at their disposal that uses propositions and hence solves the problem of unspecified relations between the concepts under study: the Implicit Relational Assessment Procedure (IRAP, Barnes-Holmes et al. 2010). Notwithstanding multiple successful applications, the IRAP’s procedure is highly complex. As a result, some participants simply fail to complete the experiment correctly (De Houwer et al. 2015). Comparatively, the RRT – although, admittedly, itself rather complicated for participants to understand – offers a simpler procedure and may in that respect be more useful with certain groups of participants than the IRAP. Nonetheless, the complexity of RRT may still prevent it from being used with some groups of participant who may struggle with the instructions. Hence, extensive pretesting with the intended respondent group is advisable.

2. Attitudes towards Standard and Colloquial Belgian Dutch

The case study reported below, focuses on the social meaning of Dutch language variation in Belgium. The current language situation in Flanders, the Dutch speaking northern part of Belgium, can be described as a diaglossic one: a continuum ranging from the standard variety to the local dialects with in between these extremes the so-called tussentaal (literally ‘in between language’), a variety we will refer to here as Colloquial Belgian Dutch (CBD, Geeraerts and Van de Velde 2013). This diaglossia grew out of a diglossic situation with only two types of language varieties, namely the local dialects and the standard variety (Geeraerts 2017, but see Jaspers and Van Hoof 2013 for an alternative view on the earlier stratification of Dutch in Flanders). Throughout the second half of the 20th century CBD developed to fill the gap between those two varieties (see for instance Geeraerts 2017 for a discussion of this evolution). CBD is the variety of choice in informal, everyday conversations, while SBD is only used for the most formal interactions. The use of dialects is reserved for local and highly informal conversations and is declining in all but the westernmost peripheral areas of Flanders (Ghyselen 2016). The CBD variety, if we can call it a variety at all (cf. Plevoets 2008), has received increasing scholarly attention in the last two decades. It is a collection of linguistic features which spans the space in the diaglossic continuum between the standard variety and the local dialects. These CBD features are situated on all linguistic levels ranging from a regional accent to phonetic, lexical and morphosyntactic variants. Not all features are equally colloquial and together with frequency, they determine the degree of CBD of a discourse (Geeraerts and Van de Velde 2013). An extensive discussion of the current language situation in Flanders can be found in for instance Grondelaers et al. (2016).

Although only a small amount of attitudinal studies about Dutch language variation have been carried out recently, their results seem to point in the direction of a tripartite structure that matches the three layers in the stratificational continuum of language use in Flanders (Rosseel et al. forthcoming). Older (i.e. 1970s and 1980s) language attitude research in Flanders indicates that dialects are judged suitable only for informal and homely situations, and, in some studies, they elicit evaluations of social attractiveness (Grondelaers 2013). Recent research does not tend to focus on attitudes towards dialectal variation, but based on findings reported in Vandekerckhove and Cuvelier (2007), Impe and Speelman (2007) and Lybaert (2014), we hypothesize that the local dialects have retained their indexical meaning of informality, friendliness and homeliness. The standard variety, by
contrast, is perceived as prestigious, signalling status and competence (e.g. Impe and Speelman 2007; Van Gijssel et al. 2008; Grondelaers and Speelman 2013; Grondelaers and Van Hout 2016). CBD is proposed to carry indexical meanings of dynamism, which is defined as a sort of cool associated with the media, and characterizes speakers as trendy and assertive. Of the studies mentioned above, Impe and Speelman (2007) is the only recent research to present that reports on all three layers of social meaning (prestige, dynamism\(^2\) and homeliness) within one study. Using a verbal guise technique, they showed that the most standard guises in their experiment received prestige evaluations, while the guise that was considered more dialectal was associated with familiarity and homeliness. The guise that included the most CBD features was linked with dynamic characteristics. In our study, we aim to gather more evidence on the social meanings of SBD and CBD, and by doing so we hope to contribute to an increased insight into the current language dynamics in Flanders.

The dynamism associations with CBD reported in previous research have been defined as ‘covert’, ‘subconscious’ or ‘deep’ evaluations and, because of that, are deemed to be particularly difficult to measure using direct attitude measures, or even using traditional indirect methods (Grondelaers and Speelman 2013; Grondelaers et al. 2016). Grondelaers and Speelman (2013), for instance, relied on the speaker evaluation paradigm to measure the social meaning of CBD. In the study reported below, we attempt to provide additional evidence for the dynamic indexicality of CBD by using a social psychological implicit attitude measure: the Relational Responding Task (RRT).

The RRT shows some similarities to the speaker evaluation methods, like the MGT, yet differs from it in ways that may allow us to measure attitudes under automatic conditions. In this study, we constructed our RRT experiment so that it would be structurally as similar to matched/verbal guise experiments as possible and hence, results could be better compared to previous research. Firstly, in our study participants listened to two audio samples representing SBD and CBD, just like in the MGT. These were recorded by the same speaker, but presented to participants as coming from two different speakers. Subsequently, participants completed two RRTs (one for prestige and one for dynamism) that measured the social meanings of the two varieties under study. The stimuli in the RRTs were statements that linked either prestige or dynamism traits to both language varieties (e.g. ‘[SBD] sounds more intelligent than [CBD]’). These traits were presented by adjectives, as is usually the case in MGT experiments.

The statements in our RRTs were formulated in a relative manner (e.g. ‘[SBD] sounds more intelligent than [CBD]’ rather than ‘[SBD] sounds intelligent’). This relative phrasing, combined with the inclusion of two varieties in the RRTs, rather than one, has a double motivation. Firstly, speaker evaluation experiments usually include multiple languages or varieties, as was the case in previous studies on the social meaning of SBD and CBD (Impe and Speelman 2007; Vandekerckhove and Cuvelier 2007; Grondelaers and Speelman 2013). If we want to be able to compare our findings to those studies, our study should be structurally and conceptually similar (Payne et al. 2008; Hofmann et al. 2005: 1380). So if previous work relied on within subject experimental designs where participants judge multiple varieties, our study should take a similar relative approach. Secondly, we believe that the evaluation of language is an inherently relative process (cf. Van Bree 1988: 3). Language varieties are evaluated in comparison to other varieties that the speaker could have used, but decided not to. Hence, making explicit which concepts are compared in the experiment ensures a better controlled study.

\(^2\) The paper originally refers to a ‘social attractiveness’ dimension, although it can be interpreted post hoc as what recent studies have termed ‘dynamism’.
Despite structuring our RRT experiment as much as possible as a traditional speaker evaluation experiment, there are of course many differences between the two approaches. One way in which they diverge is the precise measuring procedure. Whereas a MGT experiment usually features ratings on semantic differential scales or Likert scales, the RRT requires participants to categorise statements according to specific instructions. Respondents indicate whether statements are true or false. Yet, they do not react to the statements according to their own opinions, but they follow a rule that is specified in the instructions of the RRT. For instance, participants react to the statements as if they believe one language variety sounds more prestigious/dynamic than the other, rather than judging the statements based on what they actually believe about the varieties. What is measured is whether participants carry out the classification tasks correctly and how fast they perform them depending on the rule specified in the instructions. This procedure is designed with a view to measure participants’ reactions under automatic conditions. This brings us to a second point on which the RRT and the speaker evaluation paradigm differ. Linguists have used the latter to measure what they refer to as ‘covert’, ‘deep’ or ‘private’ attitudes. Although the concept is rarely discussed in much detail, covertness seems to be mainly defined in terms of unawareness: respondents are kept unaware of the fact that the researcher is trying to measure their language attitudes. The RRT, by contrast, aims to measure attitudes under conditions of automaticity, which means that it tries to measure fast and undeliberate evaluations. This may mean that participants are unaware of what is being measured, but it is not necessarily the case. Implicitness is a multifaceted concept of which awareness is only one aspect. In the RRT, for instance, it is unlikely that participants do not realize what the experiment is about, but they are not intentionally offering their personal beliefs about the attitude object (see Kristiansen & Pharao, this issue, on whether it is important to mask awareness of the attitude object in language attitudes research). Perhaps working with methods that aim for implicitness in the social psychological sense rather than covertness in the narrower sociolinguistic interpretation will allow us to gain new insights in the nature of dynamism beliefs about CBD and offer new possibilities to tap into the indexical meaning of this language variety. This study aims to take a first step towards exploring this possibility, but in any case, more research is needed to compare the status of the constructs measured by the RRT and MGT-like methods and further evaluate the potential of this new method (cf. Section 5).

3. Methodology

3.1 Design and procedure

Each participant completed two RRTs: one that measured prestige related beliefs towards SBD and CBD and one that focused on the perception of both varieties in terms of dynamism. Half of the respondents completed the prestige RRT first, the other half started the experiment with the dynamism RRT (cf. Table 1).

<table>
<thead>
<tr>
<th>Order RRTs</th>
<th># participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>prestige - dynamism</td>
<td>205</td>
</tr>
<tr>
<td>dynamism - prestige</td>
<td>186</td>
</tr>
</tbody>
</table>
Before starting the first RRT, participants were asked to listen to two sound samples representing SBD and CBD. They were instructed that what was said was not important. Instead, they had to imagine what the two persons they heard were like and keep that in mind throughout the study. Once participants had listened to both audio samples, they were allowed to proceed with the two RRTs. The precise structure of both the prestige RRT and dynamism RRT is summarised in Table 2 and Table 3 respectively. Following De Houwer et al. (2015), the order in which the rules for categorisation were presented to participants was kept constant in order to reduce additional variance in the data.4

A difficulty inherent to the RRT is that the attitude objects (in our case SBD and CBD) need to be labelled, as one has to be able to refer to them in the statements used as stimuli in the RRT. As shown in previous linguistic attitude research (e.g. Bishop et al. 2005; Coupland and Bishop 2007), the choice of a suitable label is difficult, yet crucial as it can influence the results of the attitude measurement. In the RRTs reported in this paper, we created five different conditions (between subject) in which we used five different ways of labelling SBD and CBD: explicit language labels common among non-linguists, both accurately (AN ‘General Dutch’ and tussentaal5 ‘in between language’) and non-accurately (AN ‘General Dutch’ and dialect ‘dialect’) describing the varieties, generic first names to refer to the speakers (Pieter and Jonas), names of well-known TV presenters (Bart Schols and Otto-Jan Ham) and labels referring to the sound samples themselves (fragment 1 ‘sample 1’ and fragment 2 ‘sample 2’). The results varied in interesting ways between these conditions, which shows that the labels chosen in the RRT impact the way participants conceptualize the attitude objects under study which in turn activates (slightly) different beliefs and associations. However, it is beyond the scope of this paper to report on the results of the individual conditions in detail. In what follows, we will aggregate over the outcome of the five conditions. Despite the loss of granularity this entails, the general conclusions from the RRT study do not change because of it.

---

3 Note that participants were told they heard two different speakers, even though this was not the case: they heard two clips produced by the same speaker. While preparing the experiment, we tested on a small sample of students whether this claim about two speakers was convincing. We found no evidence to suggest that it was not.

4 Future research on the RRT should look into the impact of the order in which the rules are presented to participants.

5 Tussentaal is not necessarily a label that all non-linguists in Flanders are familiar with. Yet, a pretest, as well as a post-experiment follow-up question showed the participants in this condition of our experiment were all familiar with the term.
<table>
<thead>
<tr>
<th>block</th>
<th>type of block</th>
<th>Task</th>
<th># trials</th>
<th>stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>practice</td>
<td>categorising inducer trials</td>
<td>10</td>
<td>synonyms true/false</td>
</tr>
<tr>
<td>2</td>
<td>practice</td>
<td>categorising statements according to rule 1: [SBD] sounds more prestigious than [CBD]</td>
<td>20</td>
<td>statements about prestige</td>
</tr>
<tr>
<td>3</td>
<td>experimental</td>
<td>categorising inducer stimuli and statements according to rule 1</td>
<td>20</td>
<td>synonyms true/false + statements about prestige</td>
</tr>
<tr>
<td>4</td>
<td>experimental</td>
<td>categorising statements according to rule 2: [CBD] sounds more prestigious than [SBD]</td>
<td>20</td>
<td>statements about prestige</td>
</tr>
<tr>
<td>5</td>
<td>practice</td>
<td>categorising inducer stimuli and statements according to rule 2</td>
<td>20</td>
<td>synonyms true/false + statements about prestige</td>
</tr>
<tr>
<td>6</td>
<td>experimental</td>
<td>categorising inducer stimuli and statements according to rule 2</td>
<td>20</td>
<td>statements about prestige</td>
</tr>
<tr>
<td>7</td>
<td>experimental</td>
<td></td>
<td>20</td>
<td>statements about prestige</td>
</tr>
</tbody>
</table>
### Table 3 - Structure of the dynamism RRT

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

### 3.2 Participants

In total 430 participants completed the online experiment. The link to the experiment was distributed through the online learning platform of two Flemish universities, as well as through word of mouth. Participation was voluntary and respondents were not paid. Data from 39 participants were excluded from the analyses for one of three reasons. Firstly, participants who did not meet the demographic requirements for the study (i.e. they were not originally from Flanders or Dutch was not their native language) were removed from the dataset. Secondly, participants who reported considerable disruptions while they were taking the RRT were left out as well (e.g. someone entering the room while the participant was completing the experiment). The third criterion for exclusion was related to the response latencies and number of errors in the RRT. Participants with more than 10% fast responses (i.e. < 300 ms) and those whose number of errors was more than 2.5 standard deviations higher than the mean number of errors were not included in the analyses reported below (De Houwer et al. 2015; Greenwald et al. 2003). Analyses were thus carried out on a sample of 391 participants (see Table 1).
3.3 Materials

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

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

In this study, a Brabantic variety of CBD was used. Some linguists believe that Brabantic features are spreading to other regional varieties of CBD (e.g. Goossens 2000; Taeldeman 2008). Considering this potential Brabantic expansion together with indications from previous studies that the Brabantic variety may be the most susceptible for dynamism associations (Impe and Speelman 2007), we deemed that CBD with a Brabantic flavour would give us the best chance of measuring dynamism evaluations.

The statements used in the RRTs were based on a number of traits that represent prestige and dynamism (see Appendix). To come to a selection of traits, we started out with a set of adjectives used in previous research (Grondelaers and Speelman 2013), supplemented by adjectives we selected ourselves. In a next step, pretests were carried out on a small group of university students (N = 19) to validate these adjectives and verify whether any relevant aspects of prestige and dynamism were overlooked. The inducer stimuli consisted of 10 synonyms of ‘true’ and ‘false’ and were taken from De Houwer et al. (2015) (see Appendix).

---

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

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

8 Citing Willemins (1979), Impe and Speelman (2007) explain their finding that the central Brabantic CBD variety is the only variety linked to dynamism-like traits by suggesting that peripheral varieties of CBD, like CBD with a West-Flemish flavour, are more easily perceived by listeners as dialectal. As a result, these varieties of CBD have a higher chance of being associated with homeliness and intimacy, like the traditional base dialect usually are.
3.4 Analysis

Both response latencies and error information was collected. This data was used to calculate D scores employing the $D_6$ algorithm (Greenwald et al. 2003). A D score is a difference score between the experimental blocks of the RRT that takes into account individual differences in speed between participants (see Greenwald et al. 2003 for the precise algorithm). In our study, positive D scores indicate a response tendency in line with the hypotheses (i.e. SBD sounds prestigious and CBD sounds dynamic). The higher the absolute value of the D score, the stronger the effect. The scores were calculated using the ‘IAT’ package for R (Martin 2015). We followed De Houwer and colleagues (2015) in only taking into account reaction times for the statements and not the inducer trials when calculating D scores.

4. Results

All data for the RRTs taken together (regardless of their order of appearance) show no effect for beliefs about prestige, yet they do show a significant, but small effect in the unexpected direction for dynamism (see Figure 2 and the ‘all data’ rows in Table 4). Participants seem to hold the belief that SBD sounds more dynamic than CBD. That is precisely the opposite of what was hypothesized.

Table 4 - Overview of the RRT effects. A positive mean D score indicates a pattern that aligns with our hypotheses (SBD sounds more prestigious than CBD and CBD sounds more dynamic than SBD), while a negative mean score indicates the opposite.

<table>
<thead>
<tr>
<th>RRT</th>
<th>mean D score</th>
<th>t</th>
<th>df</th>
<th>$p^*$</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>all data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prestige</td>
<td>-0.01</td>
<td>-0.54</td>
<td>390</td>
<td>n.s.</td>
<td>.03</td>
</tr>
<tr>
<td>dynamism</td>
<td>-0.05</td>
<td>-2.45</td>
<td>390</td>
<td>.04</td>
<td>.12</td>
</tr>
<tr>
<td>second RRT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>prestige</td>
<td>0.09</td>
<td>4.27</td>
<td>185</td>
<td>&lt;.001</td>
<td>0.3</td>
</tr>
<tr>
<td>dynamism</td>
<td>0.06</td>
<td>2.47</td>
<td>204</td>
<td>.04</td>
<td>.17</td>
</tr>
</tbody>
</table>

* Holm corrected for multiple testing
A closer inspection of the data reveals a strong overall order effect depending on whether participants completed the prestige RRT or the dynamism RRT first. The ANOVA of a regression model of the D scores for the prestige RRT and of one for the D scores of the dynamism RRT with RRT order, condition and the interaction of both shows a significant main effect for RRT order in both cases (prestige model: $F_{\text{order}} = 30.13, p < .001$; dynamism model: $F_{\text{order}} = 36.61, p < .001$). The main effect for condition and the interaction term do not contribute to the model significantly. This order effect is visualized in Figure 3: scores in the first RRT were generally lower than those in the second RRT. We suggest this order effect can be interpreted in terms of a training effect. We believe that the RRT procedure may have been so complex for participants to understand and perform that they needed the first few blocks to get the hang of the task they were supposed to carry out. As a result they were slower in the first set of experimental blocks compared to the second set, not because of the beliefs presented in the rules, but because they had to get used to the task. This lead to negative D scores in the first RRT rather than the positive ones we expected. Only in the second RRT were participants well acquainted enough with the procedure for us to measure the effect of following rules that either aligned with their beliefs or went against those beliefs. This explanation is supported by the data from the second RRTs. When we consider the results for those second RRTs alone, a pattern emerges that comes much closer to the results we hypothesized based on previous research (see Figure 4). For the prestige RRT, we observe a significant positive effect confirming the hypothesis (see Table 4). For the dynamism RRTs, there also is an effect in the expected direction, although it is a small one (see Table 4; see Grondelaers & Van Gent, this issue, for a parallel result with respect to attitudes toward varieties of Dutch in the Netherlands).

---

9 Note that when we split up the prestige RRTs per condition, only three of the five conditions reach significance (with Holm correction).

10 When the data are examined per condition, there is a trend in the data in line with the hypothesized outcome, yet these effects do not reach significance. This may be due to the smaller sample sizes. Although we expected that it would be easier to find evidence for dynamism evaluations for CBD in an implicit measurement of beliefs than in an explicit one, the lack of significant effects for the dynamism dimension, as opposed to the prestige
In this study we used the RRT to measure the social meanings of SBD and CBD. The RRT is designed to measure beliefs on the implicit level. The measure allows to specify precisely how the attitude object is related to a certain qualification, unlike for instance the IAT. We found implicit prestige associations with SBD and implicit dynamism associations with CBD, as hypothesized. In this respect, dimension, echoes the suggestion in previous research that positive dynamism evaluations for CBD are hard to detect even when using indirect measures (Grondelaers and Speelman 2013).
we can add more evidence to previous claims about speakers of CBD being perceived as dynamic and SBD speakers as prestigious. Yet, this conclusion was only based on half of the data collected in the experiment. Due to a training effect where participants needed some time to get to grips with the task, we had to discard the results for the first of two RRTs that participants completed. It is important to stress that this interpretation of the order effect in the data as a training effect is a post hoc explanation and no firm conclusions can be based on this analysis. Further studies will have to be conducted to verify whether similar order effects show up when using the RRT. Yet, if our interpretation of the order effect is correct, this has serious consequences for the use of the RRT as a measure for the social meaning of language variation. It would mean that a RRT study on language attitudes would have to include a training phase with an additional ‘mock’ RRT that allows participants to practice and get familiar with the method, before the actual RRT could be completed. Given that the completion of an RRT easily takes 10 to 15 minutes, this is anything but desirable. However, if we can find ways to deal with these order effects, for instance by having participants take a mini version of the RRT consisting of only a few trials or by conducting the experiment in lab conditions rather than online so participants can ask the experiment leader for clarifications, the RRT may still present a promising avenue for future research into the social meaning of language variation.

In addition to further exploring order effects in the RRT, more work is needed to shed light on the precise differences between methods like the RRT that aim to measure beliefs under conditions of automaticity and traditional methods like the MGT that capitalize on awareness. To what extent do these methods yield different results? And what does that tell us about the type of construct measured in these tasks and the theoretical concepts of implicitness and covertness? Sociolinguistic attitude research is in need of comparative methodological explorations that weigh up the opportunities offered by different methods, both novel and traditional.
References


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


Kristiansen, Tore, & Nicolai Pharao. this issue.


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

Pantos, Andrew J. this issue.


## Appendix

### Stimuli used in the prestige RRT and dynamism RRT

<table>
<thead>
<tr>
<th>Inducer stimuli</th>
<th>Prestige RRT</th>
<th>Dynamism RRT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Juist, goed, correct, exact, in orde</strong> (synonyms ‘true’)</td>
<td><strong>Juist, goed, correct, exact, in orde</strong> (synonyms ‘true’)</td>
<td></td>
</tr>
<tr>
<td><strong>Mis, onjuist, incorrect, verkeerd, fout</strong> (synonyms ‘false’)</td>
<td><strong>Mis, onjuist, incorrect, verkeerd, fout</strong> (synonyms ‘false’)</td>
<td></td>
</tr>
<tr>
<td>[SBD] <em>klinkt rijker dan</em> [CBD]</td>
<td>[SBD] <em>klinkt hipper dan</em> [CBD]</td>
<td></td>
</tr>
<tr>
<td>‘[SBD] sounds richer than [CBD]’</td>
<td>‘[SBD] sounds trendier than [CBD]’</td>
<td></td>
</tr>
<tr>
<td>[SBD] <em>klinkt slimmer dan</em> [CBD]</td>
<td>[SBD] <em>klinkt entertainender dan</em> [CBD]</td>
<td></td>
</tr>
<tr>
<td>‘[SBD] sounds more clever than [CBD]’</td>
<td>‘[SBD] sounds more entertaining than [CBD]’</td>
<td></td>
</tr>
<tr>
<td>[SBD] <em>klinkt intelligenter dan</em> [CBD]</td>
<td>[SBD] <em>klinkt relaxter dan</em> [CBD]</td>
<td></td>
</tr>
<tr>
<td>‘[SBD] sounds more intelligent than [CBD]’</td>
<td>‘[SBD] sounds more relaxed than [CBD]’</td>
<td></td>
</tr>
<tr>
<td>[SBD] <em>klinkt succesvoller dan</em> [CBD]</td>
<td>[SBD] <em>klinkt populairder</em> dan [CBD]</td>
<td></td>
</tr>
<tr>
<td>‘[SBD] sounds more successful than [CBD]’</td>
<td>‘[SBD] sounds more popular* than [CBD]’</td>
<td></td>
</tr>
<tr>
<td>[SBD] <em>klinkt ernstiger dan</em> [CBD]</td>
<td>[SBD] <em>klinkt chiller dan</em> [CBD]</td>
<td></td>
</tr>
<tr>
<td>‘[SBD] sounds more serious than [CBD]’</td>
<td>‘[SBD] sounds cooler than [CBD]’</td>
<td></td>
</tr>
<tr>
<td>[CBD] <em>klinkt armer dan</em> [SBD]</td>
<td>[CBD] <em>klinkt oubolliger dan</em> [SBD]</td>
<td></td>
</tr>
<tr>
<td>‘[CBD] sounds poorer than [SBD]’</td>
<td>‘[CBD] sounds more outdated than [SBD]’</td>
<td></td>
</tr>
<tr>
<td>[CBD] <em>klinkt dommer dan</em> [SBD]</td>
<td>[CBD] <em>klinkt saaier dan</em> [SBD]</td>
<td></td>
</tr>
<tr>
<td>‘[CBD] sounds more stupid than [SBD]’</td>
<td>‘[CBD] sounds more boring than [SBD]’</td>
<td></td>
</tr>
<tr>
<td>[CBD] <em>klinkt onwetender dan</em> [SBD]</td>
<td>[CBD] <em>klinkt gestrester dan</em> [SBD]</td>
<td></td>
</tr>
<tr>
<td>‘[CBD] sounds more ignorant than [SBD]’</td>
<td>‘[CBD] sounds more stressed out than [SBD]’</td>
<td></td>
</tr>
<tr>
<td>[CBD] <em>klinkt onbeduidender dan</em> [SBD]</td>
<td>[CBD] <em>klinkt minder populair</em> dan [SBD]</td>
<td></td>
</tr>
<tr>
<td>‘[CBD] sounds more insignificant than [SBD]’</td>
<td>‘[CBD] sounds less popular* than [SBD]’</td>
<td></td>
</tr>
<tr>
<td>[CBD] <em>klinkt zorgelozer dan</em> [SBD]</td>
<td>[CBD] <em>klinkt gereserveerder dan</em> [SBD]</td>
<td></td>
</tr>
<tr>
<td>‘[CBD] sounds more carefree than [SBD]’</td>
<td>‘[CBD] sounds more reserved than [SBD]’</td>
<td></td>
</tr>
<tr>
<td>[SBD] <em>klinkt armer dan</em> [CBD]</td>
<td>[SBD] <em>klinkt oubolliger dan</em> [CBD]</td>
<td></td>
</tr>
<tr>
<td>‘[SBD] sounds poorer than [CBD]’</td>
<td>‘[SBD] sounds more outdated than [CBD]’</td>
<td></td>
</tr>
<tr>
<td>[SBD] <em>klinkt dommer dan</em> [CBD]</td>
<td>[SBD] <em>klinkt saaier dan</em> [CBD]</td>
<td></td>
</tr>
<tr>
<td>‘[SBD] sounds more stupid than [CBD]’</td>
<td>‘[SBD] sounds more boring than [CBD]’</td>
<td></td>
</tr>
</tbody>
</table>
[SBD] klinkt onvetender dan [CBD]  [SBD] klinkt gestrester dan [CBD]
‘[SBD] sounds more ignorant than [CBD]’  ‘[SBD] sounds more stressed out than [CBD]’
[SBD] klinkt onbeduidender dan [CBD]  [SBD] klinkt minder populair* dan [CBD]
‘[SBD] sounds more insignificant than [CBD]’  ‘[SBD] sounds less popular* than [CBD]’
[SBD] klinkt zorgelozer dan [CBD]  [SBD] klinkt gereserveerder dan [CBD]
‘[SBD] sounds more carefree than [CBD]’  ‘[SBD] sounds more reserved than [CBD]’
[CBD] klinkt rijker dan [SBD]  [CBD] klinkt hipper dan [SBD]
‘[CBD] sounds richer than [SBD]’  ‘[CBD] sounds trendier than [SBD]’
[CBD] klinkt slimmer dan [SBD]  [CBD] klinkt entertainender dan [SBD]
‘[CBD] sounds more clever than [SBD]’  ‘[CBD] sounds more entertaining than [SBD]’
[CBD] klinkt intelligenter dan [SBD]  [CBD] klinkt relaxter dan [SBD]
‘[CBD] sounds intelligent than [SBD]’  ‘[CBD] sounds more relaxed than [SBD]’
[CBD] klinkt succesvoller dan [SBD]  [CBD] klinkt populairder* dan [SBD]
‘[CBD] sounds more successful than [SBD]’  ‘[CBD] sounds more popular* than [SBD]’
[CBD] klinkt ernstiger dan [SBD]  [CBD] klinkt chiller dan [SBD]
‘[CBD] sounds more grave than [SBD]’  ‘[CBD] sounds cooler than [SBD]’

* The adjective ‘popular’ was specified in the instructions to be understood as ‘having many friends’.

** In the actual RRTs, the [SBD] and [CBC] slots in the statements were filled with the different labels used for the varieties in the five conditions of the experiment (see footnote 3)