'To my surprise, I don't particularly like my own opinions': Exploring Adaptations of the 'Open-Guise' Technique to Raise Sociolinguistic Language Awareness

Mats Deutschmann (Örebro University)

Anders Steinvall (Umeå University)

Abstract
The following study describes a data-driven learning scenario aimed at raising sociolinguistic awareness of matters related to gender, language and stereotyping. The design is inspired by the matched-guise technique (MGT), a quantitative data driven experimental method that has been used extensively to investigate language attitudes. In the scenario, differences in respondents' response patterns to two gender-manipulated versions (male-female vs. female-male dyads) of the same recorded dialogue were used as a starting point for awareness-raising activities aimed at highlighting how gender stereotypes may affect perceptions of a dialogue. The main focus of the article is a comparison of the learning outcomes of two variants of the setup: a traditional undisclosed MGT-inspired setup, where the design and purpose of the experiment was kept secret until after the response phase, and a so-called open-guise design, where respondents were informed of the design and purpose of the experiment prior to the response phase. Preliminary results suggest that respondents adjust their assessments of a speaker depending on the guise, even when they know it is the same speaker they are listening to. Moreover, the open-guise design seemed to lead to greater pedagogic impact than the scenario based on the undisclosed design. However, further studies are needed to confirm these findings.

Keywords: Open-Guise Technique; Matched-Guise Technique (MGT); language attitudes; sociolinguistics; gender

1. Introduction
Pérez-Paredes et al. (2019: 145) define data-driven learning (DDL) as 'a learner-focused approach which promotes language learners' discovery of

linguistic patterns of use and meaning by examining extensive samples of attested uses of language'. According to Gilquin (2021: 231), DDL is advantageous since learning can be 'inductive and implicit', and based on exposure to authentic data. In other words, DDL affords learners' independent discovery of linguistic patterns.

To date, DDL has primarily been associated with the use of corpus data in the language classroom, a practice which began in the 1960s with the advent of corpus linguistics (McEnery and Wilson 1997). Since then, a multitude of language corpora have been used in a multitude of ways in order to afford different aspects of language learning/learning about languages (see Leńko-Szymańska and Boulton 2015 for some examples). Learning designs, however, have mainly been focussed on features related to language production, for example, tangible language aspects of grammar, syntax, and lexicon (Boulton and Cobb 2017: 380). Using corpus-based DDL in learning designs targeting aspects of sociopragmatic language competence/self-awareness related to reception, perception and language attitudes, however, is arguably more challenging. Consequently, this study presents a complement to corpus-based DDL for inductive learning targeting the learning objectives above. More specifically, this article explores the pedagogic potential of the matched-guise technique (hereafter MGT), a quantitative data-driven experimental method that has been used extensively to investigate language attitudes (see Lambert et al. 1960).

In what follows, we will give a brief overview and some critical reflections of some of the models that we have developed over the past few years in the project *Raising Awareness through Virtual Experiencing* (RAVE, funded by the Swedish Research Council), where we have used MGT-inspired designs to raise learners' sociolinguistic language awareness by exposing them to their own language attitudes and stereotypes. The article also includes a description and critical evaluation of recent adaptations of the models, where we use a so-called *open-guise* inspired design (Soukup 2013b).

2. Background

Our methods are inspired by the MGT, and later digital developments of this method that have opened up new possibilities (see Connor 2008, for example). Here, we will give an overview of MGT, some of the voiced critique of the method, as well as subsequent adaptations in response to

this critique. We will also describe our own pedagogic adaptations of MGT and summarise in what ways our methods build on/differ from the MGT.

2.1 The Matched Guise Technique

The MGT is a sociolinguistic experimental design that was initially developed to measure attitudes towards a specific language, dialect, or accent. The method has been described as having 'a neat and rigorous design aimed at people's private attitudes' (Garrett 2010: 57), and has been used in a plethora of studies, particularly in the fields of sociolinguistics and social psychology (see Garrett 2010 for a comprehensive overview).

In an MGT setup, the same text, normally a reading of a standard text, is produced in two or more variants, where the manipulated variable is the perceived identity of the speaker as manifested through language output (social/regional accent, for example). In order to eliminate as many confounding background variables as possible, the same bilingual/bidialectal actor has traditionally been used to produce the different versions. The text versions are then played to respondents who are unaware of the real purpose and design of the experiment. The method thus represents an *indirect approach* to language attitudes (Garrett 2010). MGT designs usually include a number of 'dummy' control texts in order to 'camouflage' the target texts. The respondents are then asked to rate (on a Likert scale or a semantic differential scale) their impressions of the 'different speakers' on a number of personality characteristics (for in-depth and problematized discussions of the use of such scales in research on attitudes to language, see Garrett 2010 and Soukup 2013a). The reactions elicited by each of the linguistic guises produced by the actor are then compared. Since, arguably, the only variable that varies between the different recordings is the language/dialect/accent, differences in reaction are attributed to the respondents' attitudes towards the varieties spoken, and thus, by extension, towards the social groups with which these varieties are associated.

2.2 Critique of the MGT

There have been several areas of critique raised about the MGT (see Garrett 2010: 57–59). The first area of critique concerns the authenticity of the language stimulus and the relevance of MGT to real-life language

situations. This critique was raised early, in 1971, when Lee criticized the method for being overly artificial. Note that in traditional MGT designs, the stimulus usually consists of a read passage and thus lacks situational context (see Bradac et al. 2001). This is a direct result of the method prioritising control of background variables at the expense of authenticity. The secrecy protocol, a prerequisite for the MGT, has also contributed to the exclusion of more authentic-like language stimuli; it is unlikely that respondents would conceive almost identical versions of longer strings of spontaneous speech produced by supposedly different speakers as believable.

One way around this dilemma has been between-subject designs (see Bourhis and Giles 1976 for an early example). Controlling background variables relates to the importance of creating stable frames of reference for attitudinal studies, and is further discussed by Soukup (2015). Here she argues for implementing Hymes' (1972) SPEAKING grid of eight 'components' in order to capture the key aspects of a communicative event.[1] Keeping all such variables stable in a design is, however, challenging. The complex interrelation of situational and contextual aspects was identified by Giles and Ryan (1982: 219), who pointed out that '[t]he extent to which language variety A is or is not preferred over language variety B depends upon the situation in which the assessment is made'. The use of between-subject approaches allow for more complex stimuli, including dialogues, where contextual aspects such as speaker roles and purpose can be made more conspicuous, hence (ideally) making the frame of reference more realistic.

The second area of critique concerns what is really measured in an MGT experiment. To what degree do attitudinal responses directly or indirectly relate to the people speaking the varieties? Could it be that it is the language varieties themselves that respondents react to? In a much-quoted definition, Ryan, Giles and Sebastian (1982: 7) make no such distinction. They define *language attitudes* as 'evaluative reactions

---

[1] Hymes (1972) actually identifies 16 components that can be grouped into eight, thus creating the acronym SPEAKING: Settings, Participants, Ends, Act sequences, Keys, Instrumentalities, Norms and Genres. While Soukup (2015) suggests the use of Hymes' components in the context of mixed-method research on language attitudes, the usefulness of analysing and controlling these parameters for a stable frame of reference goes beyond the mixed-method approach.

towards different language varieties or their speakers'. Soukup (2013a), on the other hand, suggests a revised terminology. Given the importance of situational and contextual factors for the evaluation by the respondents, Soukup talks about 'social meanings of linguistic variation', rather than attitudes (2013a: 263). Closely related to the above, another point of critique of the method is that the rating scale format of the evaluations compels respondents to look for contrast where they might not normally note it, thereby risking evoking stereotypical judgements that would not actually exist in an authentic unconditioned situation (see Luhman 1990).

A third area of concern has been that, even when working with the same actress/actor, it is impossible to entirely control for all unwanted background variance, such as speed of delivery, intonation, volume, and pitch (Tsalikis et al. 1991). Moreover, in so-called *verbal guise* designs, when different speakers are used, for instance, when investigating gender (Maegaard 2005), the task of controlling these variables becomes even more challenging. Here recent developments in technology have afforded entirely new possibilities, which we have used in our designs. For example, digital cut-and-paste techniques of key phonetic signal markers have enabled the creation of digitally manipulated guises based on the same recordings (see Labov et al. 2011, for example). Similarly, digital voice pitch and timbre manipulations have enabled the simulation of masculine and effeminate versions (Levon 2007) and male and female versions of the same recording (Dennhag et al. 2019).

A final area of critique of the MGT, of particular relevance to pedagogic adaptations of the MGT method (see 2.3), concerns the ethics of the method (see also 2.4). The design presupposes secrecy as to the real purpose of the experiment, an ethical dilemma which means researchers have to 'trick' students initially. Further, a believable fake context/motivation for the experiment that does not arouse suspicion has to be invented. It is difficult to ensure that respondents have not suspected the real intentions of the design. According to Kircher 2015, such responses should be excluded from the analysis. However, as discussed and illustrated by Soukup (2013b), it is virtually impossible to know whether participants have figured out the real intentions. A lack of comments does not rule out that some participants did so. Initial secrecy also means that informed consent can only be obtained after the response phase.

In order to address some of the ethical and other dilemmas listed above, Soukup (2013b: 268) has challenged the unquestioned premise of MGT studies which holds 'that informants are to be kept ignorant of the fact that they are hearing the same speaker(s) over again using different accents, varieties, or languages'. Instead, Soukup openly informed respondents of the design and purpose of the experiment prior to listening to the text versions. According to Soukup (2013b: 281), her informants had 'no problem at all in making sense of the fact that they were hearing the same speakers twice, using different linguistic varieties', and the rating patterns mirrored findings from other traditional MGT experiments investigating the same language variants. Her results show that listeners can make sense of one and the same speaker putting on 'different "coats" of identity' (2013b: 282), and successfully and honestly contrast personal impressions of social meaning in an open design. From a constructivist perspective, this can be related to speakers' everyday experiences of how their language, dialect or style may shift in relation to how their identity or role may be performed differently depending on context or community of practice (Soukup 2013a; see, for example, Holmes and Schnurr (2006) on performing different gendered identities at work).

2.3 Adapting the MGT and Open-Guise Technique for pedagogical purposes

There has been a recent increased interest in raising awareness of language variation among the public, among teachers and in schools and higher education, as evidenced by the large number of publications reporting on different efforts made in this area. Two such (random) examples are Bündgens-Kosten (2009), on improving attitudes towards African-American Vernacular English, and Hélot et al.'s (2018) volume on language awareness activities addressing multilingualism in a European education context. While the MGT and verbal guise designs have been used extensively to measure student and teacher attitudes towards different dialects, accents and languages in educational contexts (see for example Buckingham 2014; Carrie 2017; Kim 2021), there are to our knowledge no examples of instances where the potential of the method as a *pedagogical* tool to raise linguistic self-awareness of implicit language bias effects has been exploited.

Given the growing field of implicit bias training (see for example Project Implicit 2011; Sleek 2018), this is somewhat surprising. An MGT

experiment is relatively easy to set up and can be conducted in the classroom in a matter of minutes (see Kircher 2015). The analysis of the results is straightforward and gives an indication of language attitudes directly relevant for the participating group. These findings can then be used as a point of departure for discussions, self-reflections and other awareness-raising activities. Such affordances are exploited in our designs, and results to date have convinced us this type of pedagogic adaptations of the MGT can make a significant contribution to DDL aimed at raising sociolinguistic awareness in anti-bias training (see Hakelind et al. 2022; Deutschmann and Steinvall 2020; Deutschmann et al. 2021; Deutschmann et al. 2022, for example).

2.4 Designs created under RAVE

To date, we have tested and evaluated MGT-inspired learning designs in various fields, including gender and personality psychology (Dennhag et al. 2019; Hakelind et al. 2022), gender and sociolinguistics (Lindvall-Östling et al. 2020), culturally gendered stereotypes (Deutschmann and Steinvall 2020; Deutschmann et al. 2021), stereotyping of accented students in English (Lindvall-Östling et al. 2020) and in Swedish (Deutschmann et al. 2022), to mention a few. Note, that our ambitions in these activities have not been to measure language bias and attitudes as such, but rather to raise awareness of these phenomena. When necessary, pedagogical impact has thus been prioritised, rather than accurate and controlled experimental design. For example, this explains why we have not included control stimuli in our designs, as this would risk response fatigue.

In our designs, there are many aspects that are similar to previous MGT setups, but there are also differences: to date, we have followed traditional MGT secrecy protocol and students have not been fully aware of the design, or purpose, of the experiment until the debriefing phase. We have used a between-subject design (cf. Bourhis and Giles 1976), which has allowed us to use complex language stimuli generally consisting of contextualised dialogues (cf. Giles and Ryan 1982). To our knowledge no other MGT studies have done this. The recordings have been digitally manipulated in various ways depending on what language aspect is under scrutiny (cf. Labov et al. 2011 and Levon 2007; see 2.2). In our response questionnaires we have generally used fixed Likert scale rated responses,

but the more recent designs have also included an open text question of how the speakers in the recordings were perceived (cf. Luhman 1990).

The awareness-raising activities in our designs have followed similar procedures. In short, they consist of five phases: exposure to the case; response to the case; a debriefing session; a discussion seminar; and a short written reflection/evaluation (Lindvall-Östling et al. 2019; Deutschmann and Steinvall 2020). The exposure to the case and the ensuing response phase generally take place online, where students listen to a version of a recording and immediately after give their impressions on the language production of one of the characters in an online survey tool. For this phase, students have been divided into two groups (note that this is not the case in the open design described in this article: see 4.2) by the teacher or by using a digital randomizer in an online survey tool. The two groups are given access to different versions of the text.

The data from the response questionnaires is then summarized and analysed, and then presented in a debriefing seminar, where the real design and purpose of the exercise has been revealed. The debriefing is immediately followed by a discussion seminar in which students are first asked to discuss the results, their impressions, and the implications of the results in smaller groups. After this follows a class discussion where each group is given an opportunity to share and discuss their reflections with the rest of the class. Finally, we ask them to write short reflective comments on a few questions we have prepared in an online questionnaire. Although successful with regard to stimulating discussion and reflections, the above undisclosed designs have not been without issues.

Issues related to the initial secrecy protocol has created a number of challenges. In order to create maximum effect at the debriefing/discussion seminar—an 'aha moment'—keeping students unaware of the real purpose of the task has been seen as important (arguably unnecessarily so). Accordingly, we have had to make up fake reasons for including the exercise in the course. The design of the awareness-raising activities has thus to some extent been dictated by the ambition to keep the true purpose hidden from the students. Furthermore, the secrecy aspect has meant that respondents should not discuss the learning experience with potential future respondents, peers in the year below, for example. This has of course been difficult to ensure.

Another pedagogical drawback has been the fact that in any class, only half of the students listen to one of the two configuration versions of the

recording. The pedagogic impact of a within-subject design, which illustrates to an individual how his/her own specific impressions may be influenced by aspects such as accent, is thereby lost. Although playing both versions of the recording to all respondents in the debriefing session and cross-pairing of students for the debriefing discussions have been effective, experience tells us that respondents primarily relate to the version they themselves have listened to initially. In exposing students to both versions in the response phase, they have the opportunity to reflect on their own initial reactions, especially when the purpose of the setup is clear to them.

The final issue concerns ethics. From an ethical point of view, informed consent can only be obtained after the debriefing, when students are aware of the full picture. Since informed consent could not be sought in conjunction with the response phase, the data from respondents that are absent at the debriefing, or who do not answer the post-survey, cannot be used in the undisclosed guise setup. This has led to considerable data loss. Note that ethical approval was sought and gained from the Swedish Ethics Review Board,[2] who emphasised that 'informed consent' meant explaining the exact nature of the experiment.

3. Aims

The main ambition of this article is to contribute to the development of pedagogical DDL tools for raising awareness of matters related to language bias and stereotyping. More specifically, we aim to describe and critically evaluate recent developments under the project RAVE whereby we test a so-called open-guise setup (Soukup 2013b). With this ambition in mind, we will compare evaluations of the learning experiences of two groups of students who participated in similar scenario setups. The first group participated in an undisclosed MGT-inspired design, where the real purpose of the exercise was kept secret until after the response phase. Here respondents only listened to one of the two manipulations. The other group took part in an open-guise inspired design, where the purpose and design of the exercise was fully explained beforehand, and where respondents listened and responded to both manipulated versions of the recording.

---

[2] Decision 2016/75-31Ö

4. Method and material
4.1 Script and recordings
The theme of the activity in the design is communication and leadership, and how gender stereotypes may influence our impressions of a communicative event. The speech sample is inspired by dialogues described in Holmes (2005), which explores aspects related to gender and leadership. The stimulus dialogue consists of a workplace interchange where a boss (Robin) tells off an employee (Kim) for not doing his/her job properly:

> **Robin**: I assume this sort of stuff is backed up on the secure internal server, right?
> **Kim**: Eeerm. I'm… I'm not sure.
> **Robin**: What do you mean 'you're not sure'?!
> **Kim**: Well, eerm, I mean John and Beth are the ones that are involved with security and back-ups so…
> **Robin**: So if they weren't here we'd be totally lost, right… and you wouldn't have a clue!?
> **Kim**: I'd most probably look up the formal internal routines for this sort of thing… that don't exist…
> **Robin**: Well… Jesus! You're telling me you don't know, or worse, that there are no routines—this is a critical issue, don't you think? If we lose this type of stuff, or, just imagine if it ends up in the wrong hands! We are talking major disaster! Things can't be run like this!
> **Kim**: No, I guess not. Sorry, I'll try to look into it.
> **Robin**: Don't try Kim! Just do it! Give me an overview of the routines when you're done.

The script was recorded on separate tracks using a female actress enacting both of the characters, Robin and Kim. Initial audio editing was made using the software *Twisted Wave* (https://twistedwave.com) to remove unwanted pauses and sounds. The voice quality of the recordings was then manipulated using the software *Melodyne*[3], a professional post-editing tool used in the music industry. The main tool used within *Melodyne* was Pitch transition (also known as *auto tuning*). Initially, the whole monologue was

---

[3] https://www.celemony.com/en/melodyne/what-is-melodyne.

pitch transitioned. Manual pitching was then used to adjust parts that sounded unnatural in an attempt to minimize the risk that listeners perceived the audio as manipulated. It is, however, near impossible to produce a perfectly natural-sounding manipulated voice. At the end of manipulation, we had two different feminine sounding voices and two different masculine sounding voices, all of which had been manipulated. These were combined in different ways producing a total of four setups: M-F (Robin-male: Kim-female); F-M; M-M; F-F[4]. Thus, all the contextual and situational components were held constant except for the perception of the gender configuration of the speaker dyad, and any difference in the participants' responses could be related to expectations linked to gender roles.

The recordings were 'packaged' in YouTube videos depicting male/female silhouette figures engaged in a conversation in an office environment (see Figure 1).
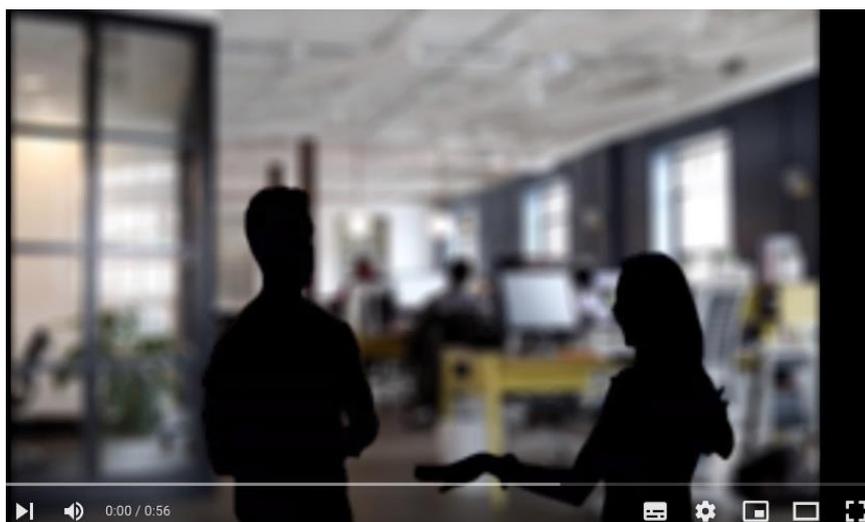

Figure 1. Example of image background to recordings (F-M version)

---

[4] The recordings are available on our website: https://www.stereotyping.se.

4.2 The response questionnaire

After the production phase, the videos were embedded in online questionnaires (SurveyMonkey), where they constituted the response stimuli. In the questionnaires, participants were first asked to give their spontaneous free text reactions to the dialogue with focus on the communicative styles and characters of Robin and Kim.

This was followed by a set of statements exploring different characteristics of Robin (the boss) and/or Kim (the employee). Respondents were asked to agree or disagree with these statements on a 7-point Likert scale, where 1 represented complete disagreement and 7 complete agreement. They were formulated as positive statements according to the format *Robin/Kim is* + adjective or descriptive phrase, and statements included descriptions relating to leadership and communicative aspects positioned on the dimensions of competence (for example, *Robin is ... effective*, *a good leader*, *straightforward and clear* etc.) and sympathy/warmth (for example, *friendly*, *abusive*, *rude*; see Figures 4 and 5 in 4.4).

In the undisclosed guise setup, a randomising tool (available in SurveyMonkey) decided which version respondents got to listen to (M-F or F-M). In the open-guise setup, respondents listened and responded to both versions. The participants were also given the opportunity to comment on any other aspects of the exercise, and in the open-guise version they were also asked to give consent (or not) that their responses be added to the research database. This was unfortunately not possible in the undisclosed guise design, as respondents were unaware of the real purpose of the exercise. Informed consent thus had to be postponed to the post-debriefing survey (something which also led to considerable loss of data: see 4.3 and 4.4).

For practical reasons, we have so far only trialled comparative scenarios exploring differences in impressions of mixed-sex setups (F-M vs. M-F). In addition, the primary focus of the quantitative analysis (statement responses) so far has been on the leader (Robin). Of course, various other comparative combinations are possible, including a comparison of all four versions with focus on both Robin and Kim. Based on previous experience from the project (see Lindvall-Östling et al. 2019), however, we have found that trying to include too many aspects in one scenario lessens the pedagogical impact of the exercise and causes confusion.

4.3 Pilot trials—participants, procedures and course context

The descriptions and data here are based on two pilot runs of the scenario that we did with classes of English student teacher trainees studying a course in sociolinguistics in autumn 2020 and autumn 2021. The pilot runs were framed as voluntary workshops. They were not part of an examination and were conducted (including the debriefing seminars) by the first author, who was not otherwise involved in the course. The rationale for choosing a course in sociolinguistics was the need for students not only to study what other studies have shown, but also to reflect on how they themselves may react to gendered linguistic behaviour. Given their subject at hand, the debriefing seminar would also give them ample opportunity to apply models and terminology on their own data. Note that given the Covid-19 restrictions that were operating at the time, all response and discussion activities were carried out online, in electronic surveys (SurveyMonkey) and an online conferencing tool (Zoom).

One group (N=30) did the exercise as a traditional undisclosed guise setup. They were thus not informed of the real design and topic focus prior to the experiment. Instead, they were told that the workshop was about aspects related to 'language, power and leadership', which was only part of the truth. This group were randomly assigned either the M-F version or the F-M version of the recording. In the second trial (N=19), we used an open-guise design and told the participants about the purpose of the exercise prior to the experiment. Participants were given access to both versions (M-F and F-M in a counterbalanced design), but were told to wait at least a day between the listening occasions in order to minimise interference of the previous impressions. In the open-guise trials, participants were also asked to give informed consent that their responses be added to our research database at this stage.

In both pilot runs, the debriefing seminars, where we revealed the design (relevant for the undisclosed guise trial only) and presented the response patterns, took place in Zoom. The discussion seminar that immediately followed the debriefing took place in so-called breakout rooms (3-4 participants in each group) followed by a whole class discussion. Finally, participants were asked to give their reflections of the learning experience in a post-survey. It was also here that the undisclosed design participants were asked to give their informed consent that the data generated by the entire trial be added to the research database. Unfortunately, relatively few respondents from this group (14 of 30)

answered the post-survey. In the open-guise trial 15 of 19 participants completed the post-survey.

4.4 Debriefing material

In this subsection, we give examples of the type of material that was presented to the participants in the debriefings of the pilot trials of this case scenario. The material is taken from the response patterns generated in the open-guise pilot trial. As such, it constitutes a summary of the group responses to the recordings and aims to illustrate how responses systematically differed (or not) depending on the version of the recording. The debriefing material from the MGT-inspired trial is not presented here since consent to publish this material is missing from 16 of the 30 participants. The response patterns from the MGT-inspired trial were, however, very similar in character to those generated in the open-guise trial. Note also that given the fact that the focus of this article is on pedagogical design and the students' learning experience, we have not included any detailed quantitative statistical comparative analysis of the responses of the two trials.

Firstly, the qualitative descriptions in the questionnaires were summarised and visualized using a world cloud tool (worditout.com). This was done by extracting all the descriptive phrases and adjectives that occurred in the free text descriptions of the characters Robin and Kim (N=19) for each guise version and saving these in separate text files for each version of the recording (M-F and F-M). These text files were subsequently used to generate word clouds (see Figures 2 and 3). The word cloud software creates an image, where the size of the words is indicative of how frequently they occur in the text.

In both pilot runs, it was evident that, although there was considerable overlap, the descriptions of male Robin were more negative and emotionally oriented than those of female Robin (see Figure 2). For example, Robin in the male guise was described as *rude* five times (vs. only once for female guise), and as *condescending* four times (once as female). In contrast, adjectives such as *concerned* (five times; none for male guise) *upset* (five times; none for male guise) and *tired* (5 times; once for male guise), which somehow explained Robin's behaviour as an understandable reaction to Kim's incompetence, were most frequent in the descriptions of the female guise (see Figure 2).
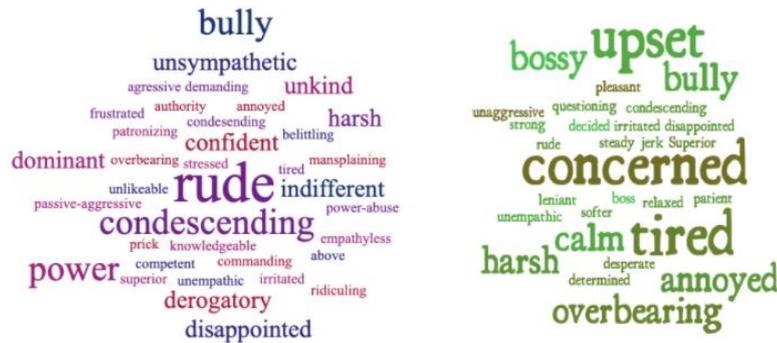
Figure 2. Debriefing material showing word clouds of adjectives and descriptive phrases used to describe Robin (the boss) after listening to the male guise (purple-blue) and the female guise (green). The largest words appeared in five of the nineteen descriptions, and the smallest once.

There were differences in the descriptions of the female and male versions of Kim too, again with obvious overlaps. Female Kim descriptions were dominated by adjectives such as *insecure* (5; 3 for male guise), and *nervous* (3; none for male guise) (see Figure 3). This semantic field also occurred in the descriptions of male Kim, but here negative descriptions inferring lack of engagement (*lazy*, *careless* and *blasé)* also occurred. Also worth noting is that the most frequent adjective used to describe male Kim was *weak* (5; 0 for female guise).
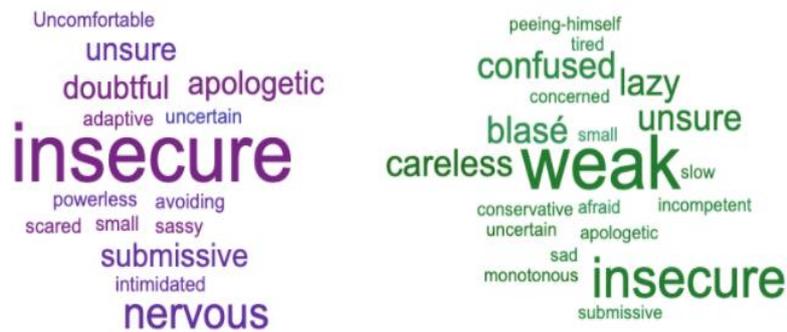

Figure 3. Word clouds of adjectives and descriptive phrases used to describe Kim (the subordinate) after listening to the female guise (purple-blue) and the mail guise (green) presented to respondents during the debriefing

The quantitative data based on the statement responses mirrored the findings from the free text data above. In the responses to the male guise, there were stronger tendencies to agree with statements referring to negative characteristics such as being *rude*, *insensitive*, *arrogant*, *aggressive*, or a *bully*. In contrast, the female version was ranked more positively with reference to professional aspects such as *competence*, being a *good leader* and being *effective* (see Figure 4).
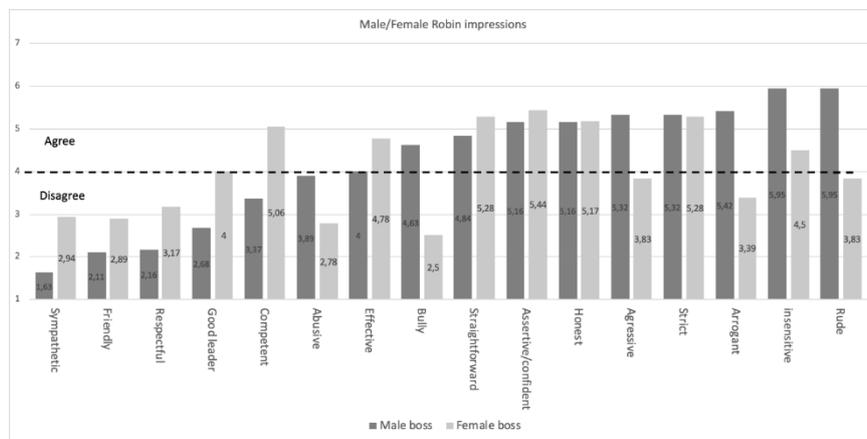

Figure 4. Debriefing materials showing impressions of traits of the boss: 'male Robin' (dark bars) and 'female Robin' (light bars). Note that 4 represents a neutral (neither nor) alternative on the Likert scale.

The largest differences in ratings between the male and female versions of the boss Robin were observed for the traits *competence*, *being a good leader* and *being sympathetic*, where the female version received more favourable ratings. For the traits *arrogance*, *rudeness* and *being a bully*, the male version received higher ratings.
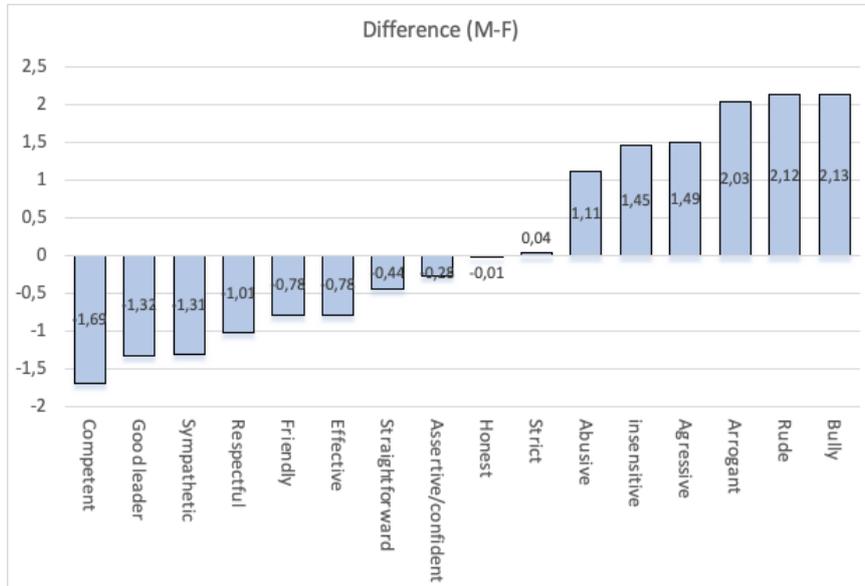
Figure 5. Debriefing materials showing differences between 'male Robin' and 'female Robin' ratings (M-F) in order of magnitude starting with negative values where 'female Robin' was rated more favourably than 'male Robin'

Figure 5 depicts the debriefing slide used to summarise the mean differences (male values - female values) between ratings of the male and female versions of the recording. Negative values imply that the male version was rated lower on the traits and positive values imply that it was rated higher. These images, then, were used as stimuli for the ensuing discussion seminar.

5. Results—comparing evaluations and self-reflections of the undisclosed guise group (UG) and the open-guise group (OG)

The results in this section are based on the evaluations of 14 respondents who participated in the secrecy protocol design (hereafter UG) trial and 15 respondents who took part in the open-guise trial (hereafter OG). The results include a summary of responses to the following questions in the post-survey aimed at capturing participants' impressions and reflections of the learning experience after the debriefing and discussion seminars:

1. Did the activity you just participated in give you any new insights? If so, what were they?
2. In what way can the experiences you gained from this experiment help you in your (future) profession?
3. Was there anything in the design of this activity that you feel worked particularly well, or alternatively that worked less well and you feel that we should change?

In addition, the survey included a 'General Comment' where respondents could leave any other reflections or comments.

5.1 Question 1—New Insights

The most common answer (8/14) in the UG group was that the exercise had not led to any new insights. In the OG group the ratio was 4/15. Many of these responses were simply formulated as short negative answers (*No*; *No, not really*; *Not really*), while others provided more comprehensive explanations that pointed to the fact that they already knew about these stereotypes but that a reminder was good, as illustrated in comment (1).

> (1) Nothing inherently new. This mostly proves the many hypotheses and theories that we've learned from before. (UG respondent)

These types of comments arguably suggest that some students at least might have been focussing on aspects of language production rather than on aspects related to perception. This points to a general danger with using this kind of setup: observed differences in how the dialogue versions are perceived may be confused with, and remembered as, differences related to production. In other words, the tendency of respondents to perceive the male guise as harsher than the female guise may mistakenly have been interpreted as an example and confirmation of the stereotype that males generally are more confident and assertive and females more submissive (cf. Cuddy et al. 2008). Some may have thought that the male and female versions of the recording actually did differ to reflect these believed structural differences. Although many responses show that participants have understood the setup, some of the comments suggest that the focus of the exercise (i.e., that it is differences in perception that we measure) needs to be particularly emphasised, especially given that this type of misunderstanding has been observed in debriefing discussions in the past

(see Lindvall-Östling et al. 2019). Arguably, listening and responding to both versions, as was the case in the OG setup, minimises this type of potential misunderstanding, as illustrated in comment (2), where the topic of the exercise is clearly signalled.

> (2) Mostly, I feel like it worked as a reminder of something I already knew, but often forget/don't think about—further proving that this is something worth being reminded of (that we're influenced by stereotypical preconceptions). (OG respondent)

Many respondents did, however, refer to new insights, more specifically 6/14 from the UG group and 11/14 from the OG group. Most of these responses were self-reflections relating to how the masculine version was evaluated more negatively: see comments (3) and (4).

> (3) I focused more on power than gender, but I would still say that I have stereotypical preconceptions concerning gender, and I realise that I interpret some behaviors more or less harsh depending on whether it is a man or a woman. (UG respondent)

> (4) I've answered as truthfully as I can in this survey, and to my surprise I don't particularly like my own opinions in a few of these cases. I thought I was less biased in regards of gender stereotypes as I'm against them in general. (OG respondent)

The comments above illustrate another potential danger with the current setup. There is a risk that discussions come to reflect a general 'male victim' backlash discourse that has become popular in the current debate of gender equality. In the full class debriefings, we have emphasised challenges in separating judgements of the individual from structural patterns. With reference to comment (3), this involved acknowledging persisting social problems, and that women are still discriminated against more than men in the workplace, while at the same time pointing to the importance that each individual case still has to be judged in isolation. For teachers in particular, the ability to see beyond the structural and to recognise the individual, while at the same time being aware structural injustices, is essential and at the core of the purpose of this exercise. In the debriefing discussions the relationship between past experiences and

structural patterns, and how these may affect judgements, was a common topic that many participants reflected on. One female respondent, for example, noted that her past experience of male bosses had coloured her judgements in the trial. We believe that this type of complex insight is the chief benefit of the setup.

In summary, the OG evaluations were more positive than the UG evaluations for the first question. For example, expressions that pointed to new insights were more common among the OG group.

5.2 Question 2—Relevance of learning experience to future professional practice

Some answers to this question in the UG group (4/14) and the OG group (5/15) addressed aspects related to how the students as future teachers would use the knowledge to monitor their own behaviour in order to avoid bias in the classroom: see comments (5)–(7).

(5) It has taught me that it is important to treat students, whatever the gender, with equal respect and pedagogical approach. (UG respondent)

(6) As a teacher student this experiment helped me see what preconceptions I have of others and also think of the preconceptions others can have of me. In the future as a teacher I will have to be conscious about the judgment I have of my students and coworkers while also considering the expectations both students and coworkers will have of me. (OG respondent)

(7) I might not 'hyper-correct' how I let speakers take space in my classroom (e.g. letting boys speak too little because I expect them to speak too much etc). (OG respondent)

Note here how comment (7) again illustrates the danger that students embrace a 'male-victim backlash' world view. As emphasised above, the importance of the ability to separate structural differences from individual differences cannot be overemphasized in the debriefing discussions. There were also several examples (eight in all) of respondents saying that they would use similar setups in their own future teaching, as illustrated in comments (8) and (9).

(8) I will probably do a similar experiment in my classroom in order to create awareness of their own expectations on gender related to professionality. (OG respondent)

(9) I want to do this kind of exercise with my students! I feel like it would be a fairly simple and effective way of making students more aware of sociolinguistics and how men and women are perceived. Especially since the subject is something that, in my experience, teachers sometimes struggle to bring into their course plans, this could be an intriguing thing to do in the classroom! (OG respondent)

The fact that so many participants were inspired by the exercise and wanted to do something similar in their own teaching was encouraging.

There were also a number of answers that concerned more general aspects of sociolinguistic awareness. These reflections addressed gender and communication as well as stereotyping in general, as illustrated in comments (10) and (11).

(10) It has made me more aware of how communication between the sexes is perceived. What might be considered okay for a woman to say to a man might not be okay for a man to say to a woman, for example. (UG respondent)

(11) Personally, the experiences gained today will hopefully serve as a reminder to myself to be more aware of my own prejudices and stereotypes. And for me to check the way I speak or the way I judge others depending on their speech more regularly. (OG respondent)

Comments such as those above seem to suggest that the raised awareness goes beyond professional contexts to include more general aspects of language and stereotyping.

In summary, both groups could exemplify how the exercise would have a direct impact on their future professional approaches. The open-guise group were slightly more enthusiastic about the method, as illustrated by the fact that the majority of respondents who said that they

would like to do something similar in their own teaching came from this group.

5.3 Question 3—Methodological reflections

Six respondents in the UG group and one from the OG group mentioned the quality of the voices in the recordings as an issue that may have influenced them in their impressions of the recording: see comments (12) and (13).

> (12) I think the discussion worked well regardless of being on Zoom. I think I noticed some issues with the voice manipulation, and maybe that affected me. You could opt for having two actors, one male and one female, but then again you would not have the same exact delivery in the lines of each actor. (UG respondent)

> (13) The quality of the voices could be better. (OG respondent)

The respondent in comment (12) shows awareness of the methodological challenges and choices this kind of design involves, where both the voice quality and delivery can affect the impression of the speaker. The voice quality of manipulated voices has been the greatest challenge during the project, something which is corroborated by responses such as these. It seems, however, that voice quality becomes less of an issue when respondents are aware that the voices have been manipulated, as illustrated by the fact that only one of the OG respondents commented on voice quality. Instead, there were examples of respondents appreciating the fact that they knew they were listening to manipulated versions of the same recording: see comments (14) and (15).

> (14) I liked that I knew beforehand that the recordings were the same person and it was just manipulated. It made me think about how I perceive male and female voices. (OG respondent)

> (15) I did also think it was interesting to do the experiment while knowing that I would react to male-female and female-male interactions. I had to examine and reflect over my judgements and expectations based on the genders of the speakers. (OG respondent)

Apart from the negative comments related to voice quality, there were other points of critique raised, primarily in the UG group, who on the whole were more negative. For example, some thought that the recordings were too short, while one respondent thought that it would have been better to use a 'real' dialogue in order to increase authenticity: see comment (16).

> (16) Is there a way of using 'real' dialogues? Would make it feel more authentic, may make the result more trustworthy (some may argue the current version feels 'forced'). (UG respondent)

Using 'real' recordings would of course increase the authenticity and contextualise the recording more clearly, but there are obvious ethical and practical dilemmas involved with using authentic data for this type of exercise. Yet another respondent made the point that we may have missed important aspects of the impressions since we dictated the focus through our choice of statements:

> (17) While it is good to 'force' a choice in grading type questions, be aware that many contextual things may fly under the radar. (UG respondent)

This is not entirely accurate, however, since we did try to capture 'things that may fly under the radar' in the open statement impression question, which was deliberately placed before the 'forced' statements so that these would not affect responses.

It is also noteworthy that one respondent in the UG setup actually questioned the undisclosed guise design and suggested that we develop an open version of the experiment, whereby respondents would listen to (and evaluate) both versions fully aware of the setup: see comment (18).

> (18) How accurate are the results since we only get to listen to the conversation between either a man and a woman or a woman and a man? Could it be an idea to get everyone to listen to both recordings? (UG respondent)

However, even in the undisclosed guise design, respondents did listen to both versions during the debriefing session when the design was revealed.

Finally, there were several positive comments, especially in the OG group. Some of these positive comments also included curiosity of other potential outcomes with other gender combinations or research designs:

> (19) I think everything was perfect! But it would have been fun to see how my impression would change if I heard a recording of Kim AND Robin being male, and vice versa. Would Robin (M) be seen as aggressive if Kim also was a man? etc; (OG respondent)

In sum, the responses from the open-guise groups were more positive and there were fewer indications that the voice quality of the recordings was perceived as a problem (both groups listened to the same recordings). The final comments from both groups also included several short positive statements such as 'Thanks for the interesting lab!'.

5.4 Summary

The evaluations from the exercise show that although a fair proportion of the students did not see how the exercise led to any new insights (especially in the UG group), the vast majority, in both groups, could see how the knowledge gained was directly applicable to their future roles as teachers. Some students, especially in the open-guise group, were also inspired by the design and said they wanted to use it themselves in the future. Regarding the reflections on method, some students in the UG group felt that the voice quality of the manipulated voices was problematic. This critique was less obvious among the open-guise group.

6. Discussion

In our attempt to use an MGT-inspired technique for pedagogical purposes in sociolinguistics, we have addressed several issues relating to the use of this design. With the use of a between-subject design, where we have used digital manipulation to alter the perception of the gender configuration of the two participants in the dialogue, we hope to have set a scene with a stable frame of reference (Soukup 2015) across all variables except for the gender of the speaker. However, the 'gender switch' does not only change the value of the 'participant component' but will also affect other contextual aspects as they are interrelated and 'overlap and mesh' (Soukup

2015: 66). For example, in our case, the contextual effects of 'power' are modified by gender expectations. Thus, this design opens up for discussions related to gender in context.

While the MGT-inspired approach with a between-subject setup has worked well for our pedagogical purposes (see, for example, Hakelind et al. 2022), we have also experienced issues, as identified above (see, for example, 2.4). This motivated us to trial an open-guise approach using the same basic setup, the merits and drawbacks of which we will discuss below. Before doing so, however, it is important to point out some serious limitations in the current study that should be kept in mind when interpreting the findings.

The findings so far are based on very limited data from two trials of the case scenario described above. We need to conduct many more trials in order to establish reliable data for response patterns so that we can elucidate what potential effects the open-guise vs. secrecy protocol designs may have on perception and on the learning experience. It may be the case, for example, that expectancy factors (Feingold 1994), influence respondents to a greater degree in the open-guise design than in the undisclosed guise design. Expectancy factors, i.e., social and cultural expectations of behaviours, may influence respondents to respond 'appropriately', rather than honestly, especially when they know that the focus of the exercise is on perceptions of gender and communicative styles. If this is the case, the main purpose of the exercise, i.e., to reveal hidden stereotypes within a group, is lost. Further, the limited data does not allow us to fully explore how the order of exposure (M-F or F-M) affects response patterns. Similarly, much more data is needed to confirm similarities/differences in the evaluations of the learning experience, and to capture all potential aspects of this experience: drawing conclusions on the qualitative responses of 14 or 15 participants is precarious to say the least. To what degree, or whether at all, these kinds of setups have any long-term effect is also not studied.

Notwithstanding these limitations, the results are interesting. Our response data arguably corroborate Soukup's findings that the 'open-guise technique actually "works"' (Soukup 2013b, 281): just as she concludes, we can confirm that respondents adjust their assessments of a speaker depending on the guise, even when they know it is the same speaker they are listening to. The question why they do so, however, remains, and based on the limited data we have we cannot dismiss that expectancy factors may

influence respondents. It should be noted that there are important differences in our design compared to that of Soukup (2013b). While her respondents heard the same speaker in a different variety, our respondents will essentially experience the exact same dialogue but with the speakers' genders 'reconfigured digitally', arguably making the setup more artificial.

The main focus of this study was not, however, on the differences of perceptions of the recordings per se, but rather on how well the open-guise vs. undisclosed guise designs work as pedagogical methods to raise sociolinguistic self-awareness. Here our qualitative response data suggest that awareness of the design and the exposure to both versions increased the impact of the learning experience. We can speculate as to why this should be the case.

Firstly, it is reasonable to assume that the open-guise design creates a greater potential for self-reflection, given the fact that respondents can compare how they themselves felt about, and responded to, the two versions, and also relate this to the response patterns of the group as a whole. As pointed out earlier, one of the disadvantages of the undisclosed design we have used previously is that respondents only listen to one version of the guise, making within-subject comparisons impossible.

Secondly, it seems that voice quality issues become less of an inhibitory factor when respondents know that the voice is manipulated. This is reasonable. Not being told about the voice manipulation prior to the experiment may well lead to respondents seeing this as a primary cause for differences in interpretations (rather than gender stereotyping effects). After all, no one likes to be 'tricked' into exposing weak spots, and when this is the case, we may subconsciously try to find excuses to explain our behaviour. The 'aha moment' may be negatively coloured by the indignation of the feeling of being tricked.

Finally, we would argue that the logic behind the exercise was easier to convey in the open-guise design, thereby better preparing the students to fully understand the data presented to them and the logic of the exercise. Although not evident in this particular study, our experience from previous undisclosed design RAVE activities is that it is sometimes difficult to bring home to students the fact that they have actually been exposed to different versions of the *same* recording.

In summary, we can conclude that the results from the pilot trials discussed in this article are promising. The open-guise approach unclutters

some of the ethical and methodological dilemmas of the undisclosed design and allows us to more openly explore the multi-dimensionality of stereotyping effects. Moreover, these pilot trials also give an indication of the potential that DDL based on student data could have in subjects such as sociolinguistics. Further trials are needed, however, to assure the quality of the method.

References

Boulton, Alex, and Tom Cobb. 2017. Corpus use in language learning: A meta-analysis. *Language Learning* 67(2): 348–393. doi:10.1111/lang.12224.

Bourhis, Richard Y., and Howard Giles. 1976. The language of cooperation in Wales: A field study. *Language Sciences* 42: 13–16.

Bradac, James J., Aaron Castelan Cargile, and Jennifer S. Hallett. 2001. Language attitudes: Retrospect, conspect and prospect. In *The new handbook of language and social psychology*, edited by W. Peter Robinson and Howard Giles, 137–155. New York: John Wiley.

Buckingham, Louisa. 2014. Attitudes to English teachers' accents in the Gulf. *International Journal of Applied Linguistics* 24(1): 50–73. doi:/10.1111/ijal.12058.

Bündgens-Kosten, Judith. 2009. Teachers' attitudes toward African American Vernacular English: Influence of contact with linguistics on ambivalent attitudes. PhD diss., Rheinisch-Westfälischen Technischen Hochschule Aachen. https://publications.rwth-aachen.d e/record/51259/files/Buendgens_Kosten_Judith.pdf

Carrie, Erin. 2017. 'British is professional, American is urban': Attitudes towards English reference accents in Spain. *International Journal of Applied Linguistics* 27(2): 427–447. doi:/10.1111/ijal.12139.

Connor, Robert T. 2008. Computer simulation and new ways of creating Matched-Guise Techniques. *International Multilingual Research Journal* 2(1/2): 102–108. doi:/10.1080/19313150802010319.

Cuddy, Amy J. C., Susan T. Fiske, and Peter Glick. 2008. Warmth and competence as universal dimensions of social perception: The Stereotype Content Model and the BIAS Map. *Advances in Experimental Social Psychology* 40: 61–149. doi:/10.1016/S0065-2601(07)00002-0.

Dennhag, Inga, Anders Steinvall, Camilla Hakelind, and Mats Deutschmann. 2019. Exploring gender stereotypes about interpersonal

behavior and personality factors using digital Matched-Guise Techniques. *Social Behavior and Personality* 47(8): 1–13. doi:/10.2224/sbp.8150.

Deutschmann, Mats, and Anders Steinvall. 2020. Combatting linguistic stereotyping and prejudice by evoking stereotypes. *Open Linguistics* 6(1): 651–671. doi:/10.1515/opli-2020-0036.

Deutschmann, Mats, Eric Borgström, Daroon Yassin Falk, Anders Steinvall, and Johan Svensson. 2022. 'It ain't what you say it's the way that you say it': Using Matched Guise Technique (MGT) to raise awareness of accent stereotyping effects among Swedish pre-service teachers. *Language Awareness.* doi:/10.1080/09658416.2022.2067556.

Deutschmann, Mats, Anders Steinvall, and Mattias Lindvall-Östling, 2021. Raising awareness about gender and language among teacher-training students: A cross-cultural approach. *Open Linguistics* 7(1): 666–684. doi:/10.1515/opli-2020-0181.

Feingold, Alan. 1994. Gender differences in personality: A meta-analysis. *Psychological Bulletin* 116(3): 429–456. doi:/10.1037/0033-2909.116.3.429.

Garrett, Peter. 2010. *Attitudes to language*. Cambridge: Cambridge University Press.

Giles, Howard, and Ellen Bouchard Ryan. 1982. Prolegomena for developing a social psychological theory of language attitudes. In *Attitudes towards language variation: Social and applied contexts*, edited by Ellen B. Ryan and Howard Giles, 208–223. London: Edward Arnold.

Gilquin, Gaëtanelle. 2021. Using corpora to foster L2 construction learning: A data-driven learning experiment. *International Journal of Applied Linguistics* 31(2): 229–247. doi:/10.1111/ijal.12317.

Hakelind, Camilla, Anders Steinvall, and Mats Deutschmann. 2022. The power of aha! On stimulating and guiding students towards self-awareness and critical reflection while teaching about personality psychology and gender stereotypes. *Psychology Learning & Teaching* 21(1): 57–72. doi:/10.1177/1475725720979460.

Hélot, Christine, Carolien Frijns, Koen van Gorp, and Sven Sierens (eds.). 2018. *Language awareness in multilingual classrooms in Europe: From theory to practice.* Berlin: De Gruyter.

Holmes, Janet. 2005. Leadership talk: How do leaders 'do mentoring', and is gender relevant? *Journal of Pragmatics* 37(11): 1779–1800. doi:/10.1016/j.pragma.2005.02.013.

Holmes, Janet, and Stephanie Schnurr. 2006. 'Doing femininity' at work: More than just relational practice. *Journal of Sociolinguistics* 10(1): 31–51. doi:/10.1111/j.1360-6441.2006.00316.x.

Kim, Sugene. 2021. English as a lingua franca in Japan: Multilingual postgraduate students' attitudes towards English accents. *Journal of Multilingual and Multicultural Development.* doi:/10.1080/ 01434632.2021.1909053.

Kircher, Ruth. 2015. The Matched-Guise Technique. In *Research methods in intercultural communication*, edited by Zhu Hua, 196–211. John Wiley. doi:/10.1002/9781119166283.ch13.

Labov, William, Sharon Ash, Maya Ravindranath, Tracey Weldon, Maciej Baranowski, and Naomi Nagy. 2011. Properties of the sociolinguistic monitor. *Journal of Sociolinguistics* 15(4): 431–463. doi:/10.1111/j.1467-9841.2011.00504.x.

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

Lee, Richard R. 1971. Dialect perception: A critical review and re-evaluation. *Quarterly Journal of Speech* 57(4): 410–417. doi:/10.1080/00335637109383086.

Leńko-Szymańska, Agnieszka, and Alex Boulton (eds.). 2015. *Multiple affordances of language corpora for data-driven learning.* Amsterdam: John Benjamins.

Levon, Erez. 2007. Sexuality in context: Variation and the sociolinguistic perception of identity. *Language in Society* 36(4): 533–554. doi:/10.1017/S0047404507070431.

Lindvall-Östling, Mattias, Mats Deutschmann, and Anders Steinvall. 2019. 'Oh it was a woman! Had I known I would have reacted otherwise!': Developing digital methods to switch identity-related properties in order to reveal linguistic stereotyping. In *Virtual Sites as learning spaces: Critical issues on languaging research in changing eduscapes,* edited by Sangeeta Bagga-Gupta, Giulia Messina Dahlberg, and Ylva Lindberg, 207–239. England: Palgrave Macmillan. doi:/10.1007/978-3-030-26929-6_8.

Lindvall-Östling, Mattias, Mats Deutschmann, and Anders Steinvall. 2020. An exploratory study on linguistic gender stereotypes and their effects on perception. *Open Linguistics* 6(1): 567–583. doi: /10.1515/opli-2020-0033.

Lindvall-Östling, Mattias, Mats Deutschmann, Anders Steinvall, and Satish Patel. 2020. 'That's not proper English!': Using cross-cultural matched-guise experiments to raise teacher/teacher-trainees' awareness of attitudes surrounding inner and outer circle English accents. *Educare* 3: 109–141. doi:/10.24834/educare.2020.3.4.

Luhman, Reid. 1990. Appalachian English stereotypes: Language attitudes in Kentucky. *Language in Society* 19(3): 331–348. doi:/10.1017/S0047404500014548.

Maegaard, Marie. 2005. Language attitudes, norm and gender: A presentation of the method and results from a language attitude study. *Acta Linguistica Hafniensia* 37(1): 55–80. doi:/10.1080/03740463.2005.10416083.

McEnery, Tony, and Andrew Wilson. 1997. Teaching and language corpora (TALC). *ReCALL* 9(1): 5–14.

Pérez-Paredes, Pascual, Carlos Ordoñana Guillamón, Julie Van de Vyver, Alice Meurice, Pilar Aguado Jiménez, Grainne Conole, and Purificación Sánchez Hernández. 2019. Mobile data-driven language learning: Affordances and learners' perception. *System* 84: 145–159. doi:/10.1016/j.system.2019.06.009.

Project Implicit. 2011. Project Implicit. Accessed 20 October 2021. https://implicit.harvard.edu/implicit/.

Ryan, Ellen Bouchard, Howard Giles, and Richard J. Sebastian. 1982. An integrative perspective for the study of attitudes towards language variation. In *Attitudes towards language variation: Social and applied contexts*, edited by Ellen B. Ryan and Howard Giles, 1–19. London: Edward Arnold.

Sleek, Scott. 2018. The bias beneath: Two decades of measuring implicit associations. *APS Observer* 31(2): 11–14.

Soukup, Barbara. 2013a. The measurement of 'language attitudes': A reappraisal from a constructionist perspective. In *Language (de)standardisation in Late Modern Europe: Experimental studies*, edited by Tore Kristiansen and Stefan Grondelaers, 251–266. Oslo: Novus.

Soukup, Barbara. 2013b. On matching speaker (dis)guises: Revisiting a methodological tradition. In *Language (de)standardisation in Late Modern Europe: Experimental studies*, edited by Tore Kristiansen and Stefan Grondelaers, 267–285. Oslo: Novus.

Soukup, Barbara. 2015. Mixing methods in the study of language attitudes: Theory and application. In *Responses to language varieties: Variability, processes and outcomes*, edited by Alexei Prikhodkine and Dennis R. Preston, 54–84. Amsterdam: John Benjamins. doi:/10.1075/impact.39.03sou.

Tsalikis, John, Oscar W. deShields, and Michael S. LaTour. 1991. The role of accent on the credibility and effectiveness of the salesperson. *Journal of Personal Selling & Sales Management* 11(1): 31–41.