Language comprehension under uncertainty: The pragmatic implications of non-native accents and the role of individual differences

by

Veranika Puhacheuskaya

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics
University of Alberta

Abstract

This dissertation explores the pragmatic implications of speaking with a non-native accent when the listeners or readers are native speakers. Specifically, it focuses on the comprehension of language associated with either linguistic uncertainty, such as irony and metaphors, or epistemic uncertainty, such as statements of uncertain truth. As non-nativeness is commonly associated with less accuracy in semantic choices, reduced shared knowledge and lower social status, as well as disrupted processing fluency, it should have negative consequences for the perception of uncertain language. On the other hand, literal language or facts known to be true or false may remain unaffected. Further, it explores whether comprehension of non-native speech is modulated by individual differences, offering the first investigation of generalized implicit foreign accent bias as a predictor of non-native language processing while also examining the effect of explicit attitudes toward non-native accents, political orientation, and cognitive and personality traits.

This dissertation comprises four behavioural experiments and one acoustic analysis. The first two experiments investigate linguistic uncertainty in different modalities, while the other two examine epistemic uncertainty in different modalities. Experiment 1 investigates the perception of native and non-native irony using dialogs with natural prosody. The results indicate that all types of non-native irony are perceived as less ironic and cause greater hesitation during interpretation. This effect is unlikely to stem from reduced intelligibility, as demonstrated by both lack of correlation between intelligibility and irony ratings and subsequent work in the field with converging findings that used no oral speech (Bazzi et al., 2022). While political leaning significantly predicted irony comprehension accuracy, this effect was uniform across native and non-native speakers. An ad hoc fine-grained acoustic analysis of the stimuli used in Experiment 1 suggested that non-native speakers may engage in compensatory behaviour wherein they prosodically mark all irony types significantly more than native speakers and even invest extra prosodic effort when the social cost of being misunderstood is too high, as is the case with ironic praise. Experiment 2 investigates the perception of metaphorical language in a strictly written modality. The results suggest that all sentences are perceived as less sensical when attributed to an immigrant speaker with a strong non-native accent. Incorporating the non-native speaker identity also takes more cognitive effort, as reflected in longer processing and evaluation times. Multiple cognitive, social, and personality

variables modulated the results. Experiments 3 and 4 investigate whether non-nativeness affects truth evaluation of domain-specific unknown facts. While substantial work in the field has investigated this issue before, the results remain extremely inconclusive (e.g., Foucart et al., 2020; Lev-Ari & Keysar, 2010; Souza & Markman, 2013). Competing theoretical frameworks attribute the potential lower credibility of non-native speakers to either processing disfluency or social categorization. The present experiments focus on domain-specific facts and probe the malleability of social categorization effects by manipulating the group membership of non-native speakers, introducing them as either Canadian citizens or immigrants. Experiment 3 uses written modality and shows weak evidence that non-native speakers may be considered less credible on unknown statements. Experiment 4 uses auditory modality and, surprisingly, shows that people are more inclined to believe *all* statements spoken by a non-native individual introduced as a Canadian citizen. This suggests a modality-specific group membership change that brings about a "knowledgeable immigrant" effect. This effect was limited to monolingual participants and absent for bilinguals. In addition, Experiments 2 to 4 employ an Implicit Association Test with knowledgeability attributes, which reveled a significant bias against non-native speakers. Crucially, the strength of this bias modulated truth ratings in Experiment 4.

Together, the findings of this dissertation highlight the powerful influence of social expectations in language comprehension and show that the effects of non-nativeness extend beyond processing fluency, interacting with personal and social factors. Based on the findings of this dissertation, I argue that the linguistic identity of the speaker manifested in their accent has a larger negative impact on the resolution of linguistic uncertainty than epistemic uncertainty, but the effects of non-nativeness are not always limited to a particular trope and may affect figurative and literal sentences alike.

Preface

Examining committee:

Juhani Järvikivi, Supervisor
Anja Arnhold, Supervisory Committee
Sidney J. Segalowitz, Supervisory Committee
Herb Colston, University Examiner
Alice Foucart, External Examiner

All studies reported in this dissertation received research ethics approval from the University of Alberta Research Ethics Board 2: project *English Conversations*, Pro00102750, 10 September 2020 (Chapter 2); project *Not so Stubborn Things*, 5 September 2022 (Chapter 5), and project *Does this sentence make sense?*, Pro00134190, 12 September 2023 (Chapter 4). Recruitment for all the experiments was supported through a Social Sciences and Humanities Research Council (SSHRC) Partnership Grant, "Words in the World," 895-2016-1008, awarded to Dr. Järvikivi.

The research reported in this dissertation was completed in collaboration with Dr. Juhani Järvikivi who assisted in conceptualization, methodology, and data analysis. Dr. Anja Arnhold assisted in conceptualization and data analysis for Chapter 3. Dr. Järvikivi, Dr. Arnhold, and Dr. Segalowitz provided feedback in the writing process and assisted in the editing of the manuscript. I was responsible for experiment design, data collection, data analysis, and manuscript writing. Undergraduate assistants Claire Weidemann, Gwen McGuire, Elizabeth Wood, Maureen Luckwell, Taylor Lundstrom, and Victoria Blair assisted with EEG data collection for experiments reported in Chapter 5 (EEG data itself is not reported here).

Contents of Chapter 2 were published as:

Puhacheuskaya, V., & Järvikivi, J. (2022). *I was being sarcastic!*: The effect of foreign accent and political ideology on irony (mis)understanding. *Acta Psychologica*, *222*, 103479. https://doi.org/10.1016/j.actpsy.2021.103479

Puhacheuskaya, V., & Järvikivi, J. (2021). Oh, the irony!: Interpersonal variation in the processing of foreign-accented and native irony. *Proceedings of the Annual Meeting of the Cognitive Science Society,* Vol. 43.

Contents of Chapter 4 were published as:

Puhacheuskaya, V., & Järvikivi, J. (2025). Sorry, you make less sense to me: The effect of non-native speaker status on metaphor processing. *Acta Psychologica*, *254*, 104853. https://doi.org/10.1016/j.actpsy.2025.104853

Contents of Chapters 3 and 5 have not been published yet.

Acknowledgements

There are not enough words in all the languages I speak—and, as is well known, the sky is the limit for linguists—to express proper gratitude to my doctoral supervisor, Juhani Järvikivi. I was very nervous about coming to Canada for this degree, in no small part due to my impostor syndrome working overtime, but you have been so incredibly supportive and encouraging from the very first day we met that it somehow tricked me into believing that *maybe* I belonged here. Your unwavering optimism and generosity with your time made this journey not only possible but also tremendously enjoyable. Thank you for always explaining the most difficult concepts in the clearest possible way and for always finding the right words to reignite my motivation in those low "Keep your jack!" moments—which were plenty. Changing the entire topic of the dissertation due to a one-in-a-lifetime pandemic was no joke. Thank you for giving me so much independence but at the same time always being there as a mentor, a guide, a friend, and a constant source of inspiration. I am going to miss dearly our "quick weekly check-ins" where we inevitably invent a new form of time dilation, but I am looking forward to the next chapter in our collaboration.

I am also deeply grateful to Anja Arnhold and Sid Segalowitz for being on my committee. Anja, your work ethic is something else. If there was a Perfectionist Hall of Fame, your name would adorn the main wall—although I am sure you would have a few comments on how to improve the plaque. It has been a pleasure to work with you, truly. Sid, both the width and the depth of your knowledge on any random subject is incomprehensible. I could not have wished for a better introduction to all things EEG, and your seminars gave me so much food for thought that I am still chewing on it, years later. It has been an immense privilege to have you on my committee, and I hope our paths cross again in the future.

I would like to extend special gratitude to Alice Foucart and Herbert Colston for reviewing this dissertation. Your insights and unfathomable expertise on the topic were extremely helpful, and I could not have wished for better reviewers. Thank you for making my Ph.D. defence really enjoyable. Special thanks to Lila Daskalaki for chairing it all and for being so extremely supportive over the years.

I am grateful to each and every member of the linguistic department for making it feel like home for these five years and for the occasional much-needed distraction from school. It has been a blast.

Thank you to my undergraduate assistants Taylor Lundstrom, Claire Weidemann, Gwen McGuire, Elizabeth Wood, Maureen Luckwell, and Victoria Blair for their help with experiments. Taylor, you made EEG collection so much more enjoyable, and I am fairly certain I would have lost my sanity without you. Your optimism is so strong it illuminates the entire building. I am also grateful to Lindsay Griener for her patience with my constant memory lapses and for conducting the CCP lab orchestra so smoothly.

I also want to thank Joanna Błaszczak for encouraging my initial interest in psycholinguistics and bringing me into the world of academia. Joanna, you are the reason I am here today. You may no longer be with us, but you will stay in my heart forever. Special thanks to Anna Czypionka for everything you have done for me. I am immensely grateful.

Outside of the insurmountable walls of academia, I am forever grateful to my two best friends, Anna Gurskaya and Anna Padkavyrava. Thank you for providing the perfect balance of dark Eastern European sarcasm and genuine around-the-clock support that were both fundamental in preserving my fragile grip on sanity. You are the best. Dziakuj.

I am, of course, very grateful to my family for their support as well as never-ending questions about what I am going to do with my life after graduation (to the surprise of absolutely no one, the answer is still “I don’t know”). Dad, you always were and still are my role model in life. Thank you for igniting my love to science, cosmos, and especially birds.

It would only be appropriate to thank my two cats, Luna and Dante, for being my emotional support animals and for brightening those endless hours of work from home (and for their help with typing this dissertation, albeit with questionable success).

And, of course, my deepest gratitude goes to my partner, Chris, for all his love and unconditional support over these tremendously long roller-coaster years. Thank you for so gracefully carrying the burden that comes with having an exhausted, sleep-deprived graduate student as a life partner, and for always reminding me that I am capable of anything and being so proud of my achievements. You are the reason this dissertation is finally done.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

A large body of research from social psychology has convincingly demonstrated that a person's linguistic identity, as manifested in their standard accent (i.e. used by the majority of speakers where communication takes place) or non-standard accent (i.e. non-native[1] or used by minorities in that region), profoundly affects social evaluations of the speaker (see Fuertes et al., 2012 for a comprehensive meta-analysis). At the same time, the slowly mounting psycholinguistic research has started to show that online processing of non-native speech differs significantly from that of native speech (e.g., Caffarra et al., 2019; Caffarra & Martin, 2019; Foucart et al., 2019; Grey & van Hell, 2017; Hanulíková et al., 2012; Romero-Rivas et al., 2015, 2016). Most of this research has focused on syntactic errors and semantic anomalies, with studies reporting attenuated syntactic processing with less surprisal and reanalysis and modulated semantic processing (albeit not consistently). Fewer studies, however, have investigated the pragmatic implications of speaking with a non-native accent when the language under consideration is syntactically correct but characterized by interpretative uncertainty–specifically, figurative language and statements of a generally unknown truth value. Whereas the nature of uncertainty in figurative language is linguistic (i.e. participants need to disambiguate between literal and ironic/metaphoric interpretation of the sentence[2]), the uncertainty in unknown facts such as "Ants don't sleep" is epistemic (i.e. most participants lack the necessary world knowledge to evaluate their truth value). Understanding such language thus requires an active interpretative process that is likely to take the

[1] Most commonly, the accent of non-native speakers is referred to as a "foreign" accent. In monolingual regions, "foreign" and "non-native" accents are effectively synonymous. However, in many multicultural societies, such as the Canadian one, the distinction is not that simple. Native speakers of Quebec French using English as a second/third language have a non-native accent that is not "foreign" as they are domestic speakers. While native speakers might perceive domestic non-native accents differently from immigrant accents, research in bilingual Belgium suggests that accent similarity still drives ingroup/outgroup categorization, with people being extra careful not to accidentally include a wrong member into their ingroup (Yzerbyt et al., 1995). The accents used in this dissertation are Chinese and Southern European, both being foreign to North America where experiments took place, hence foreign/non-native distinction is less relevant. See section 1.2 for the definition and discussion of the terms "native" and "non-native".

[2] Although the mere existence of "literal" language dichotomous with "figurative" language has been debated (see Gibbs, 1984; Giora, 2002).

identity of the speaker into account. For instance, to infer the non-literal meaning, listeners/readers might consider wider communicative context, speaker's tone of voice and other actual or simply inferred information about the speaker. Similarly, for statements whose truth value is either hard or impossible to evaluate, they may engage in speculation to assess the credibility of the given information using other cues (Schwarz, 2004), which again is likely to involve the assessment of the speaker's knowledge. Since non-native speech is associated with at least three distinct factors, such as disrupted processing fluency due to its segmental and suprasegmental deviations from native speech (e.g., Anderson-Hsieh et al., 1992, 1992; Gut, 2012; Hanulíková & Weber, 2012; Munro & Derwing, 1995; Wester et al., 2007), expectation of errors due to still developing LX[3] command (e.g., Gibson et al., 2017; Hanulíková et al., 2012; Lev-Ari & Keysar, 2012), and lower social status due to social categorization into an outgroup (e.g., Paladino & Mazzurega, 2020; Pietraszewski & Schwartz, 2014; Rakić et al., 2011), I hypothesise that a non-native accent will have pragmatic implications during the processing and interpretation of language characterised by interpretative uncertainty but not literal language or known world facts.

The main goal of this dissertation is thus to investigate whether a non-native accent has pragmatic implications for the perception of figurative language as well as statements of uncertain truth. The second objective of this dissertation is to investigate the amount of interpersonal variation in such implications. Any conversation is an event that is shared between the speaker and *the listener* (Fraser & Kelly, 2012). Recent trends in psycholinguistics have seen a shift in handling individual differences, slowly moving away from the idea of an invariant cognitive architecture for language where any interpersonal variation is discarded as noise (Kidd et al., 2018). Although this view facilitated the development of population-based models of language production and comprehension (Cutler & Clifton, 1999; Friederici, 2012; Grice, 1975), these one-size-fits-all models unfairly assume less individual variation than is generally the case between healthy adults. More and more studies show the effect of individual differences even on core online language machinery (e.g., Cichocka et al., 2016; Eekhof et al., 2021; Hubert Lyall, 2019; Hubert Lyall & Järvikivi, 2021; Van Berkum et al., 2009a). Despite the fact that research on individual differences in non-native speech processing and interpretation has been very limited and mostly confined to the segmental and word level (i.e. ability to recognize accented sounds; see Fraser & Kelly, 2012;

[3] Following Dewaele (2017), I use the term "LX" to refer to any language acquired after L1, of any proficiency level.

Yi et al., 2013, 2014, among others), and to a very narrow range of individual differences, mostly global cognitive abilities (short-term and working memory, attention, vocabulary knowledge, or familiarity with the accent; Janse & Adank, 2012; Porretta et al., 2016, 2020), it can be reasonably hypothesized that pragmatic implications of speaking with a non-native accent would be subject to individual variation. After all, correlational studies in psychology assume the existence of meaningful personality-based differences for numerous areas of life (Soto, 2019, 2021), so language interpretation should not be any different.

This introductory chapter is structured as follows. I first briefly discuss the pragmatics of sentence comprehension and how extralinguistic information shapes "the pragmatic meaning", including both real-time processing and the final outcome of this processing. I also quickly discuss what is currently known about the resolution of ambiguity associated with figurative language as compared to epistemic ambiguity associated with world knowledge. Next, I summarize previous research on the effects of non-native speaker identity on language processing and interpretation. I begin by discussing terminological problems around the terms "native" and "non-native". I then transition to discussing recent psycholinguistic findings on the effects of non-native speaker status on online language processing as well as offline linguistic judgments. Importantly, I identify and discuss three distinct factors associated with non-native speech–acoustic deviations from native speech, incomplete LX command, and social categorization into an outgroup–and how they affect language processing and interpretation, either independently or in combination. I then review converging findings from social psychology that a non-native accent has a profound and mostly negative effect on the social judgements of the speaker. I additionally discuss the issue of methodology in non-native speech research. Finally, I provide a comprehensive overview of prior research on individual differences in language processing and interpretation and the measures I focused on in this dissertation: explicit and implicit accent attitudes, political orientation, cognitive flexibility, empathy, and general language background variables. I then turn to discussing the present dissertation, including its research questions and general layout.

1.1 The pragmatics of sentence comprehension

Communication would not be possible if people could not compute the meaning of what is said and respond to that meaning. How exactly that happens, however, is far from established. To begin

with, what exactly is included in "sentence meaning" and in how many steps is it computed? Many traditional linguistic accounts assume the existence of the context-invariant, "semantic" meaning that is computed by retrieving the lexical meaning of words from the long-term memory and organizing them into hierarchical structures based on grammatical rules (Grice, 1975; Levinson, 2000; Searle, 1975, 1979, see also Fodor, 1983). These accounts are known as "The standard pragmatic model" and assume that the "semantic" meaning is always constructed first (and rapidly). It is, however, commonly agreed upon that such meaning is extremely rudimentary and incomplete, and that many more mechanisms for refining and enriching it are required (e.g., Münster & Knoeferle, 2018; Van Berkum, 2009). These refining mechanisms take into account the available *extra*linguistic information—i.e. information that is external to the linguistic meaning itself and includes details about the communicative context in which the conversation takes place, the particular speaker, and the personal values of the listener themselves—to compute the extended, "pragmatic", meaning. This meaning is traditionally assumed to be constructed second (and more slowly). Usually the term "pragmatic inference" is used to mean both the *process* during which this extended, context-dependent meaning is computed and the *outcome* of this process (Elder & Haugh, 2024). In a nutshell, pragmatics is broadly defined as inferring meaning situated in context (I. Martin & McDonald, 2003; Mey, 2001, 2003).

How correct this two-step account of sentence interpretation is and whether semantics can be treated as autonomous with respect to pragmatics is actually a matter of debate. There are both theoretical and empirical reasons to doubt its validity. From a theoretical perspective, as pointed out in Sztencel (2018), Van Berkum (2009), and Van Berkum, Van Den Brink, et al. (2008), the very existence of some context-invariant meaning is questionable, as, for instance, indexical resolution depends on the communicative situation (understanding the meaning of *I* or *you*), and even the meaning of "self-sufficient" lexical words inevitably depends on the situation and the speaker (e.g., Marrville, 2017; Niemi et al., 2020). From an empirical perspective, time-sensitive studies using methods like electroencephalography have shown that the neurophysiological response to the "semantic" meaning of a sentence is immediately affected by extralinguistic information, with many studies showing neural signatures that are indistinguishable from that of the semantic meaning composition (Bornkessel-Schlesewsky et al., 2013; Hagoort et al., 2004; Regel et al., 2011; Van Berkum, Van Den Brink, et al., 2008; Van Berkum et al., 2009b).

Moreover, fMRI research has shown that speaker-related meaning, semantic meaning, and world-knowledge meaning are computed in overlapping cortical regions (inferior frontal gyrus and left middle temporal gyrus), which points to a shared neural substrate for the computation of linguistic and extralinguistic information. As succinctly put by Gibbs & Perlman (2010), "there is no evidence that people automatically create literal, semantic, purely propositional representations for sentences (i.e. a 'sentence meaning') before elaborating on these representations to infer speakers' and writers' broader communicative messages (i.e. a 'speaker meaning')" (p. 307). That said, some studies do lend support to the competing idea that the computation of this extended meaning takes longer than computing the semantic meaning alone–for instance, Martin et al. (2014) showed that integrating general world knowledge takes about 200 ms longer than computing the "purely" semantic meaning (e.g., *bark* elicits a faster ERP response than *smoke* in a sentence like *Before the age of eight, children start to…*) (cf. Hagoort et al., 2004).

In summary, the very existence of enriched pragmatic meaning that includes speaker and listener information, as well as situational communicative context, is not debated, but the relative independence of this meaning from a more "bare-bones" semantic meaning and the order of their computation remain unclear. Importantly for our purpose, the computation of this "pragmatic" meaning may rely on resolving various types of ambiguities. To clarify this further, I will now briefly discuss the two types of ambiguities this dissertation is concerned with and the general processes associated with their resolution.

1.1.1 Linguistic uncertainty

Linguistic uncertainty can arise from multiple different sources: lexical ambiguity (e.g., *bank*) and vagueness (e.g., *soon*), structural ambiguity (e.g., saw a man *with binoculars*), referential ambiguity (e.g., Martha told Nina that *she* was wrong), and, of course, ambiguity associated with figurative language (e.g., ironies, metaphors, metonymies, idioms, etc.) (e.g., Gibbs Jr., 2002), indirect requests (e.g., Bašnáková et al., 2014), and other phenomena, like quantifier scope ambiguities (Kurtzman & MacDonald, 1993). As Macagno & Bigi (2018) point out, resolution of linguistic ambiguity is not just about deciding what a particular lexical item means, it is about establishing the best interpretation of the utterance in broadly understood *context* (see also Mey, 2001, 2003).

Figurative language and the cognitive processes involved in its comprehension have been the subject of hundreds of empirical studies and theoretical analyses over the years. Multiple theories have been proposed (see, for instance, Gibbs, 1984; Holyoak & Stamenković, 2018; Steen, 2023) and multiple factors affecting figurative language comprehension have been identified (e.g., Al-Azary & Buchanan, 2017; Arzouan et al., 2007; Caillies et al., 2019; Cardillo et al., 2012; Gibbs, 1984; Giora, 1999; González Fuente, 2017; Kowatch et al., 2013; Obert et al., 2018). As succinctly summarized by Gibbs & Colston (2012): "We maintain that it is quite unlikely that certain default processes occur apart from contextual influences in each case of interpreting figurative meaning and that people do not always use contextual information in the same way in all discourse situations. The temporal course of figurative meaning interpretation depends on numerous factors that include the types of figurative language, *the people involved, their likely goals and motivations, social and cultural context*, local discourse interactions, and (quite importantly) *the specific task* used to study how language is processed and what is understood." (p. 126; emphasis mine).

Naturally, different types of figurative language may require their own interpretation strategies. For instance, ironies and metaphors have been argued to differ in their communicative functions and processing demands. Whereas metaphors convey the similarity between what is said and what is implied, irony conveys the opposition of what is said and implied (Winner, 1993). Colston & Gibbs (2002) also argue that irony expresses the attitude of the speaker toward the situation and thus requires second-order inferences about the speaker's mental state, whereas metaphors express something new about a particular topic and thus do not require such inferences. That said, it would not be unreasonable to suggest that both irony and metaphor convey *some* aspect of the speaker's attitude toward a given subject. By using a metaphor, for instance, the speaker deliberately selects comparisons that reveal their perspective on the subject or their emotional evaluation of it. Hence, the identity of the speaker and their intentions should play a role in the comprehension of both ironies and metaphors.

Generally speaking, the correct interpretation of figurative language relies heavily on figuring out the intentions of the speaker, and listeners will use different cues to deduce that intention (Colston & Gibbs, 2002; Gibbs, 1986; Gibbs et al., 1991). For instance, the information the reader possesses about the author of a metaphor has been shown to immediately constrain and shape its

interpretation (Gibbs et al., 1991). During comprehension of ironic remarks or indirect requests, listeners have been shown to adopt, both cognitively and affectively, the perspective of the speaker (Bašnáková et al., 2014; Spotorno et al., 2012). Broadly speaking, *everything* that listeners know or infer about the speaker has been shown to be rapidly and automatically incorporated during speech processing (e.g., Bergen & Grodner, 2012; Bornkessel-Schlesewsky et al., 2013; Jiang et al., 2020; Kroczek & Gunter, 2021; Regel et al., 2010; Van Berkum, van den Brink, et al., 2008), and non-native speaker identity should not be an exception.

1.1.2 Epistemic uncertainty

Epistemic uncertainty arises from a lack of knowledge on a particular subject, which can stem from multiple different sources, such as incomplete data, future unpredictability, or human cognitive limitations. Importantly, when people encounter facts of unknown veracity, they naturally turn to other cues that can help them evaluate a particular fact as true or false (Oppenheimer, 2008; Schwarz, 2004). One such factor proposed in the literature is processing fluency defined as the subjective perception of effort associated with a mental task (Oppenheimer, 2008). Alter & Oppenheimer (2009) identify at least three types of processing fluency: conceptual fluency affected by semantic priming, perceptual fluency affected by visual clarity, and linguistic fluency affected by phonological priming or simplicity. Previous studies manipulating different kinds of fluency found that more fluent stimuli are judged as truer than less fluent ones (e.g., Begg et al., 1992; McGlone & Tofighbakhsh, 2000; Reber & Schwarz, 1999; Souza, 2012; Unkelbach, 2007).

Another factor that naturally affects the resolution of epistemic uncertainty and language comprehension in general is the perceived *speaker knowledge*, i.e. how much the listener thinks the speaker knows about a particular topic. For instance, Bergen & Grodner (2012) argued that underinformative sentences such as “Some squirrels hibernate” can be interpreted in two different ways: either this sentence implies that not all squirrels hibernate, or it indicates that the speaker is uncertain whether all squirrels hibernate. They showed that the interpretation the participant settles on is directly influenced by whether the speaker is introduced as an expert on the topic or not. Importantly, the inferences from literally conveyed meaning in their experiment was unaffected

by the manipulation. Speaker's certainty expressed in their tone of voice has also been shown to affect world knowledge integration (Jiang et al., 2020).

Naturally, credibility should differ depending on both what is known about the speaker themselves *and* the topic under discussion. For instance, knowing that someone is a welder should make them credible on matters related to metalworking, material properties, and so on, but, without further information, does not necessarily make them more credible on topics related to gardening or the political situation in the Middle East (despite subjective confidence in the opposite idea often witnessed on social media). Taking this argument a step further, one can hypothesise that adult immigrants may be expected to know less about certain topics related to the country where they immigrated and more about certain topics related to the country from where they emigrated (see Fairchild & Papafragou, 2018, p. 90 for a similar idea). This idea, though theoretically supported, has yet to be empirically tested.

In summary, the resolution of both linguistic uncertainty and epistemic uncertainty responds to a variety of factors, including the identity of the speaker. While speaker identity encompasses aspects like gender, age, and race—characteristics usually apparent from one's appearance—accent represents a less studied yet significant dimension. The following section provides a comprehensive review of how non-native accents impact various aspects of communication.

1.2 Non-native speaker status in communication

The very notion of a "native speaker" along with its idealization in psycholinguistic research (sometimes referred to as "native speakerism", see Holliday, 2006) has been recently subject to criticism and re-evaluation, with researchers questioning both its definition and utility (Davies, 2004; Kamhi-Stein, 2016; Medgyes, 1992). The situation becomes even more complex with the term "non-native speaker". Many see the concept of defining someone through what they are not as inherently problematic and advocate for terms like "L2/LX user" or "new speaker" instead (Cook, 2002; Dewaele, 2017; O'Rourke & Pujolar, 2013). Not being a native speaker is problematic not only because it is a circular definition but also because this approach reflects a broader phenomenon from social psychology where minorities are defined as merely different from something that does not need an explicit definition (Taifel, 1981). This is further complicated

by the fact that LX speakers of English, however broadly or narrowly defined, far outnumber L1 speakers of English at the moment (Eberhard et al., 2021) and thus themselves constitute a majority. Yet other researchers are dissatisfied with treating the native/non-native distinction as rigid binary categories and advocate for a more fluid approach, where speakers may "switch" categories depending on the circumstances (e.g., a highly advanced LX speaker may pass as a native speaker under proper conditions) (Piller, 2002). Outside the field of applied linguistics, however, native and non-native speakers remain omnipresent, despite the growing criticism of the terms themselves and the concepts behind them (Llurda, 2009). Much of it is attributed to the work by Chomsky, who introduced the notion of the idealized native speaker as the only reliable source of linguistic data (Chomsky, 1965).

Davies (2003) proposes to treat the native speaker status as a *social construct* instead, membership in which is defined by self-ascription, recognition as such by others, and language proficiency. This approach appears the most comprehensive. Self-ascription is a valuable tool often employed in psycholinguistics, although some researchers argue that the question "Are you a native speaker of X?" is inaccurate and needs to be replaced with more targeted and explicit questions regarding using a particular language in a particular time period (Cheng et al., 2021). Recognition as such by others also consistently serves as a reliable tool. Davies (2004) argues that native speakers "know" other native speakers of their language variety, in no small part due to shared cultural knowledge alongside shared pragmatic and paralinguistic cues. As far as accents go, Munro & Derwing (1995) have shown that listeners tend to agree on who has a native or a non-native accent with exceptional intraclass reliability. Accent is an important component of non-nativeness, since converging findings across many studies convincingly show that a native-like accent is near-impossible to achieve when the speaker is an adult (see Birdsong, 2006; Dollmann et al., 2020; Piske et al., 2001; Scovel, 2000; Tahta et al., 1981, among others). As for proficiency, Davies (2003) argues that even people who fit the first two criteria of being a native speaker still have varying levels of language proficiency (see also Abrahamsson & Hyltenstam, 2008, p. 491 for supporting findings). Non-native speakers obviously differ from each other in language proficiency, and Davies (2003) claims that achieving a native-like competence is harder when one goes past the Chomsky (1965)'s definition of it. While Chomsky pioneered the concept of linguistic competence, it is widely accepted nowadays that communicative competence as

introduced by Hymes (1972) also forms an integral part of speaker competence. Broadly defined as the knowledge of the socio-linguistic functions of language in communication (Stern, 2003), communicative competence has been argued to be much harder to acquire even by early LX learners who typically cannot "live out the cultural routines" as native speakers do and instead have to learn them more artificially (Davies, 2003, p. 115). Since native speakers recognise each other by shared cultural and pragmatic knowledge, among other things, it is reasonable to assume that reduced communicative proficiency will predict perception by native speakers as a non-native speaker.

The studies in this dissertation will focus on non-native speakers who have a distinct non-native/foreign accent but do not make grammatical mistakes (as stimuli are controlled by the experimenter), hence they do not belong to either extreme of a non-native speaker continuum discussed above. Native speakers as participants will be identified using the relatively uncontroversial definition of "acquiring language in early childhood" in all the studies.

In short, even though the categories may be difficult to define, psycholinguistic research has shown that they have measurable effects on language processing. Next section discusses these effects in more detail.

1.2.1 Disruption of processing fluency

Naturally, the acoustic signal itself makes a major contribution to the processing of non-native speech. People are exceptionally quick at detecting a non-native accent, with studies showing that as little as 30 ms of the word "two" is enough for native English listeners to reliably tell apart native English from French-accented speech (Flege, 1984). They can even tell apart native and non-native speakers in languages they *do not speak* (Major, 2007), which made the author argue that perhaps non-native speech has salient universal features.

Of primary importance to the discussion of non-native accents is the concept of processing fluency discussed above. Because non-native speech deviates from native speech on both segmental and suprasegmental levels and exhibits high variability, research has shown that cognitive processing of non-native speech is usually more effortful (Anderson-Hsieh et al., 1992, 1992; Gut, 2012; Hanulíková & Weber, 2012; Munro & Derwing, 1995; Wester et al., 2007). It is common for non-native speakers to replace LX phonemes that do not exist in their L1 with a native sound (Wester

et al., 2007) or to transfer prosodic features from their L1 to LX to mark information structure (Van Maastricht et al., 2016). The larger mismatch between the native listener's stored phonological representations and the acoustic properties of non-native compared to native speech entails increased cognitive effort, therefore listening to non-native speech has been compared to listening under adverse conditions, since it shares many characteristics with acoustically degraded speech (Van Engen & Peelle, 2014). Naturally, research has shown that adverse listening conditions disrupt processing fluency (H. Lane, 1963; Munro, 1998). Even when non-native speech is fully intelligible (i.e. listeners are able to correctly transcribe it in its entirety) it still requires more time and effort to process (Munro & Derwing, 1995). Even when non-native speech is paired with a speaker's face, it still incurs greater processing cost despite general processing facilitation (Grey et al., 2020).

Crucially, processing fluency is an important meta-cognitive cue that affects judgments across a broad range of social and linguistic dimensions, over and above the actual content of the stimulus (e.g., Oppenheimer, 2008; Souza, 2012; Winkielman & Cacioppo, 2001). Less fluent stimuli cause more negative affect (Reber et al., 1998), are judged as worse category members (Oppenheimer & Frank, 2008) and as less true (McGlone & Tofighbakhsh, 2000), and lead to harsher evaluations of the speaker (Dragojevic & Giles, 2016). Unsurprisingly, non-native speech likewise leads to harsher social evaluations of the speaker (see section 1.2.3.1 for a detailed discussion) (Dragojevic et al., 2017; Dragojevic & Giles, 2016). The strength of the accent also correlates with more negative speaker attitudes (Dragojevic et al., 2017). Since social judgments are also affected by the speaker's social status (Bradac et al., 1976; Bradac & Wisegarver, 1984), this state of affairs is further complicated by the fact that phonetic features of non-native accents, such as the substitution of interdental fricatives with plosives (Wester et al., 2007), are often shared with features found in lower class dialects of English (Ryan, 1983).

Theoretically, it could be expected that perceptual adaptation to a particular accent would improve objective processing fluency, which would, in turn, improve social evaluations of the speaker, but evidence so far disconfirms this (Ogden, 2020). Ogden came to the conclusion that subjective processing fluency affects social judgments more strongly than the objective fluency. It is, however, possible that this subjective perception is affected by the listener's life-long familiarity with an accent, and the literature provides evidence that accent familiarity matters for many

linguistic processes (e.g., Porretta et al., 2016, 2020; Witteman et al., 2013). More research on the topic is needed.

In summary, native listeners are surprisingly good at telling apart native and non-native speech on the basis of extremely short speech segments, and non-native speech is more cognitively taxing to process even when it is fully intelligible by objective measures.

1.2.2 Expectation of low(er) language proficiency

Despite the lack of a clear definition of native and non-native speakers, it is commonly agreed upon that most non-native speakers are characterised by the still developing LX command. This entails that non-native speech tends to be error-prone and less reliable in conveying the speaker's intention (Lev-Ari & Keysar, 2012). The important consequence of this is that listeners actively *adapt* to the processing of non-native speech on multiple linguistic levels. Phonetically, it has been demonstrated that native listeners are more likely to relax their vowel categories for non-native rather than native speakers (Hay et al., 2006). Purely labelling the speaker as Canadian/Michigan or Australian/New-Zealander is enough to cause changes in vowel perception, even when the listeners admit to being aware of the fakeness of this manipulation (Hay et al., 2006; Niedzielski, 1999). Semantically/pragmatically, listeners may be less likely to adopt a literal interpretation of the sentence and more likely to give the non-native speaker the benefit of the doubt. For instance, Gibson et al. (2017) showed that implausible sentences such as *The businessman benefited the tax law* are more likely to be interpreted in a plausible way when delivered with a non-native accent. Likewise, Fairchild et al. (2020) and Fairchild & Papafragou (2018) demonstrated that pragmatically infelicitous sentences such as *Some people have lungs* are perceived as more sensical when readers believe they came from non-native speakers. In terms of the very manner of processing, listeners engage in more shallow processing and are less likely to notice a word change when the speaker is non-native (Lev-Ari & Keysar, 2012). Both syntactically and semantically, listeners assume that non-native speakers are highly variable in their production and stop being surprised by their errors. Non-native errors have been shown to evoke an N400 but not the P600 effect assumed to be responsible for re-analysis processes (Hanulíková et al., 2012; Romero-Rivas et al., 2015), which is additionally modulated by the error typicality (Caffarra & Martin, 2019) and whether listeners identified the accent or not (Grey & van Hell, 2017). These studies convincingly

show that the brain responds differently to errors that are expected vs. unexpected. Some authors even argue that non-native accents reduce the width of lexical activation (Romero-Rivas et al., 2016). Importantly, since non-native speakers are perceived as grammatically unstable, it has been shown that listeners tend to judge grammatically correct sentences as ungrammatical when they are foreign-accented (Wesołek et al., 2023). The authors referred to this effect as the "illusion of ungrammaticality" or "grammatical tinnitus" wherein the mere expectation of errors in non-native speech makes listeners perceive errors when they are not there. This effect is somewhat orthogonal to the findings of Gibson et al. (2017) where listeners give non-native speakers the benefit of the doubt and perceive implausible sentences as more plausible. To reconcile these findings, one can argue that while people may hear grammatical errors that are not there, they may also realize that non-native speakers know more than they are able to communicate at this state of their LX knowledge, and thus give more lenient pragmatic judgments.

In short, a non-native accent causes processing changes on multiple levels of processing since listeners assume a higher incidence of errors and adapt to them. This can result in a diverse set of effects wherein listeners perceive non-native speech as more sensical/plausible when it is not and at the same time grammatically incorrect when it is, in fact, correct. The next section discusses an aspect of non-nativeness that is often ignored in psycholinguistic research—social categorization of the speaker.

1.2.3 Social categorization into outgroup

Humans are a socially complex species. The human frontal cortex is wired more complexly than in any other primates and plays a central role in regulating social behavior, with social categorization being a rapid and automatic feature of the human mind (Ito & Bartholow, 2009). It has been shown that the brain groups people by race, gender, and social status in the earliest stages of processing (as indicated by the N100 and P200 amplitudes, components generally associated with attentional focus) (Ito & Urland, 2003; Rule et al., 2012; Zink et al., 2008). Specifically, P200 is presumed to reflect facilitated processing of "motivationally or emotionally relevant stimuli" (Paulmann et al., 2013, p. 7). Naturally, social categorization and heightened attention to outgroups has evolutionary roots. As pointed out by Greene (2014), this inherent feature of our brains designed for tribal life promotes ingroup cooperation and outgroup distancing. Foreign-accented

speech has also been found to trigger more negative P200 amplitudes compared to native speech (Romero-Rivas et al., 2015). Interestingly, Foucart et al. (2019) reported an early negativity associated with just *looking* at the face of a non-native as compared to a native speaker. Some authors have argued that the outgroup status may even be a more influential factor than fluency in shaping perceptions of non-native speakers (Gluszek & Dovidio, 2010; Mai & Hoffmann, 2014).

How consistently non-native speakers are categorized into an outgroup is currently unclear, and factors influencing this categorization remain largely unexplored. As argued by Ryan (1983), monolingual listeners who speak standard language varieties are more likely to instantly categorize non-native speakers as an outgroup, which entails not only dissimilarity from themselves but also lower societal status. Generally speaking, however, the preference for native accents found in previous research seems to be robust. To begin with, when ethnicity and accent provide conflicting cues, adults consistently choose accent over visual ethnic cues for social categorization (Paladino & Mazzurega, 2020; Rakić et al., 2011). To quote the authors of one of the studies, "it was almost as if participants became blind to the visual category information in the presence of more meaningful auditory category information" (Rakić et al., 2011, p. 24). Studies examining the effect of such categorization showed that faces presented with native utterances are remembered and recognized better than those with non-native utterances (in line with the outgroup homogeneity effect), although fascinatingly, this effect was present for the Swedish-Spanish language pair but not the Swedish-English one (Champoux-Larsson et al., 2022). Naturally, this bias does not appear only in adulthood. Child research has shown that accent preferences precede racial preferences in the course of a child's development (Kinzler et al., 2009; Kinzler & Spelke, 2011). Because of this behaviour by very young children, some scholars have argued that accent historically was a more reliable cue to membership than race, since neighboring communities more likely differed in accents than races (K. A. Collins & Clément, 2012). Infants also prefer to look at native over non-native speakers before they can even comprehend speech, and children prefer to have native over non-native speakers as their friends (Imuta & Spence, 2020; Kinzler et al., 2007, 2009; Nazzi et al., 2000). In short, native accent preference is exceptionally strong.

More important than the social categorization itself are, of course, the consequences of that categorization. People tend to assume that ingroup members share their values, beliefs, and attitudes, and that outgroup members have a contrasting perspective (Dovidio & Gaertner, 2010).

They also tend to ascribe more diverse traits to their ingroup members, perceiving the outgroup as more homogenous (Tajfel et al., 1971). Unsurprisingly, converging findings from action perception, face perception, and information processing, as well as across multiple disciplines such as neuroscience, psychology, sociology, and political science point to persistent ingroup favoritism and/or outgroup antagonism (see Ito & Bartholow, 2009 for a review of EEG studies and Molenberghs & Louis, 2018 for a review of fMRI studies from different disciplines). Stereotypes based on the "us vs. them" dichotomy are commonly viewed as a means to structure and simplify a complex social world (Macrae & Bodenhausen, 2000) but are also at the core of prejudice in attitudes and discrimination in behavior (Hewstone et al., 2002; Neuberg & Schaller, 2008). Interestingly, countries where national identification is strongly tied to language exhibit stronger correlation between national identification and anti-immigrant prejudice according to surveys conducted in 31 countries (Pehrson et al., 2009). Given the above, it is unsurprising that non-native speakers consistently face discrimination by native speakers manifested in decision-making and levels of cooperativeness (Birney et al., 2020; Giles & Billings, 2004; Spence et al., 2024).

How malleable the outgroup categorization of non-native speakers is remains largely unknown. Research on intergroup bias has shown that social categories are fluid and can also be completely arbitrary (also called "minimal") (Diehl, 1990; Reichl, 1997; Tajfel et al., 1971; Tajfel & Turner, 1986). A neighbor next door can be an ingroup member when we play on the same volleyball team, and an outgroup member because her ethnicity is different. Another person can be part of the ingroup if they underestimate the number of dots on the screen and the outgroup if they overestimate it (Tajfel, 1970; Tajfel et al., 1971). Hence, some outgroup effects can be reduced by either activating or creating a different identity of the speaker that is shared with the listener (the so-called common in-group identity, see Gaertner et al., 1993). For instance, Van Bavel et al. (2008, 2011) showed that assigning people of different races to experimentally created groups increases fusiform face area activity for ingroup compared to outgroup faces regardless of their race. Interestingly, activating a different identity for bilingual individuals through language can already reduce prejudice. For instance, Ogunnaike et al. (2010) showed that Arabic-French bilingual participants exhibit stronger pro-Moroccan attitudes when they complete the Implicit Association Test (for a more detailed explanation of the test, see section 1.3.1) in Arabic than French, whereas Spanish-English bilinguals exhibit more pro-Hispanic attitudes when they

complete the test in Spanish rather than English. Recent research on accent perception showed that verbal labels with explicit group membership (Canadian/New Canadian/Immigrant) significantly affect the perceived accent strength. Specifically, people rated the speaker's accent as stronger when the speaker was presented with an Immigrant/New Canadian label as compared to no label, whereas a Canadian label did not affect accent ratings (Calkins et al., 2023). Ryan & Sebastian (1980) explicitly manipulated social status of the speakers and found that middle-class accented speakers were rated higher than lower-class ones on almost every dimension except for speech intelligibility. In line with these findings, Foucart et al. (2019) manipulated the social status of native and non-native speakers through auditory introductions (where the speaker self-introduced themselves as more or less successful) and then reinforced that status through a "hierarchy game" (where the successful speaker always ended up in the highest-ranking position). Although the low-ranking non-native speaker was excluded from the analysis for independent reasons, the study showed that accent and status both modulated world knowledge integration as indicated in the ERP components. The N400 and the P600 amplitudes were more negative for unknown versus true statements only for the high-status native-accented speaker, suggesting shallower processing of unknown statements attributed to the foreign-accented speaker and thus their possible lower credibility.

A counterpoint to the studies discussed above needs to be mentioned, which may suggest that the extent to which accent-based social categorization can be overridden may be somewhat limited. Specifically, Pietraszewski & Schwartz (2014) argued that accent behaves as a dedicated social dimension (akin to gender and age) rather than as a by-product of coalitional categorization (akin to race). If some factor is a by-product of coalition, people notice and remember that factor markedly less when other cues of coalitional affiliation (say, being on the same football team) no longer correspond to it. If, however, something behaves like a distinct social dimension, then people continue to robustly and implicitly encode this factor even if it is no longer predictive of affiliation. Using a memory confusion paradigm and focusing on U.S. vs. British accents, they found that categorization by accent was not reduced by this manipulation.

In summary, there is emerging evidence that explicit cues to social group membership can affect how non-native speakers and their speech are perceived. The next section describes the measurable effects of social categorization on explicit social evaluations of the speaker.

1.2.3.1 Social evaluations

As discussed above, social categorization leads to tangible behavioral and evaluative consequences. Research in social psychology has shown that interpersonal relations, including language attitudes, are organized around two main broad psychological dimensions, a vertical dimension of status/power/dominance and a horizontal dimension of solidarity/attractiveness/friendliness (Locke, 2003). Generally speaking, (socio-intellectual) status judgments include knowledge-related evaluations (intelligence, competence, knowledgeability, education) as well as evaluations related to success and social class. Solidarity judgments include friendliness, similarity to the rater, warmth, likeability, agreeableness, and kind-heartedness (Bradac & Mulac, 1984; Bradac & Wisegarver, 1984, 1984; Fuertes et al., 2012; Giles et al., 1981; Locke, 2003). These two dimensions underpin the more recent Stereotype Content model (SCM) (Cuddy et al., 2008; Fiske et al., 2002), which, in turn, is a precursor to an even more recent Behaviors From Intergroup Affect and Stereotypes model (BIAS) (Cuddy et al., 2007).

It is commonly agreed upon that such trait attribution is driven by social categorization, which in turn is driven by language features, such as accent (Weyant, 2007) or just lexical ingroup/outgroup markers (Mange et al., 2009). As far as accented speakers go, social psychology has already amassed a large body of research showing that people evaluate non-native speakers lower on at least the status dimension, but often on the solidarity dimension as well (for research that found speakers of non-standard accents to be upgraded on solidarity traits instead see, for instance, Hogg et al., 1984; Ryan & Carranza, 1975). A comprehensive meta-analysis found that speakers of non-standard (including non-native) accents are evaluated lower on both status (mean Cohen's d = 0.99, 95% CI = 0.83–1.16) and solidarity (mean Cohen's d = 0.52, 95% CI = 0.33–0.70) (Fuertes et al., 2012). The overall magnitude of the effect across all the dimensions was also large (d = 0.82). As indicated by the authors, some evaluations were shaped by literally a single word, such as "hello". The magnitude of these effects are at the upper bounds of those observed in psychology, which agrees with the findings from child and adult studies that the preference for native/standard accents is robust and trumps even race (Imuta & Spence, 2020; Kinzler et al., 2007, 2009; Nazzi et al., 2000; Rakić et al., 2011).

Summarizing this section, non-nativeness has been shown to affect both social evaluations of the speaker and linguistic processing of their speech. Importantly, the effect of accent is a multifaceted

phenomenon that not only has multiple root causes (disruption of cognitive fluency, expectation of inaccurate lexical choices and grammatical errors, as well as outgroup social categorization) but also dynamically interacts with other social and communicative variables. Next section discusses yet another factor affecting comprehension of non-native speech—personality of the *listener*.

1.3 Listener effects in non-native speech processing

Variation is ubiquitous in biological systems. Humans differ along most dimensions, from purely physiological to behavioral and emotional. Despite this being a widely accepted fact and despite the "psycho-" in psycholinguistics, the common approach historically has been to downplay the possibility of meaningful differences between healthy L1 native speakers and leave them to social psychology (Kidd, 2012). When such differences were explored, they have been mostly confined to atypical cases with established biological correlates (clinical disorders) or, in LX research, ultimate attainment in language proficiency. The situation has been a bit different in the domain of language use, which has provided extensive evidence that moral values, political beliefs, and emotions routinely affect one's linguistic expression (G. Lakoff, 2002; Mehl et al., 2006; Wu, 1995). The issue is thus that not much progress has been made in the domain of language comprehension—that is, with a focus on a listener rather than a speaker, although it is reasonable to assume that the emotional relevance of the speaker's message triggers fast and mostly unconscious evaluation and appraisal, affecting speech processing (van Berkum, 2018). It has been shown that affect and cognition are tightly intertwined in the neural architecture (Adolphs & Damasio, 2012), and hence affective appraisal cannot be separated from the core language processing machinery (van Berkum, 2018).

Moreover, the rapidly developing field of personality neuroscience has already begun to provide evidence that numerous traits, such as openness, fondness for hierarchy, stereotype susceptibility, or attributional styles not only shape how one perceives the world but also have tangible biological sources (DeYoung & Gray, 2009). As one pertinent example, liberal views have been associated with larger amounts of gray matter in the anterior cingulate cortex (ACC) involved in empathy whereas conservatism has been associated with an enlarged amygdala involved in threat perception (see Kanai et al., 2011 but also see a large recent pre-registered replication of the result by Petropoulos Petalas et al., 2024 that confirmed the amygdala finding but not the ACC finding). In

light of the above, it becomes unlikely that the language architecture would remain completely unaffected by personal traits, values, and beliefs.

As discussed in the beginning of the chapter, psycholinguistic research has recently seen a shift in its treatment of interpersonal variation. Studies using various methodologies have found that linguistic processing is affected by moral views (Van Berkum et al., 2009b), mood (B. A. Converse et al., 2008; Egidi & Caramazza, 2014; Federmeier et al., 2001; Vissers et al., 2013), political leaning (Hammond-Thrasher & Järvikivi, 2023; Hubert Lyall & Järvikivi, 2021), empathy (S. Li et al., 2014; van den Brink et al., 2012a), social orientation (Schirmer et al., 2008), personality traits (Hubert Lyall, 2019; Jiménez-Ortega et al., 2022; see also B. B. Li & Huang, 2024 for an excellent overview), and of course biological factors like sex (Schirmer et al., 2002) or age (Wlotko et al., 2012), among many others. These findings led Münster & Knoeferle (2018) to argue that at least two groups of socially relevant comprehender's characteristics need to be included in a model of language processing for it to be descriptively adequate: biological factors (such as age) and experiential factors (such as affective states, literacy, world knowledge and stereotypes).

In what follows, I will review individual variables that this thesis intends to investigate, as these variables can be theoretically linked to either non-native speech processing specifically or to the processing of figurative language and world knowledge facts.

1.3.1 Implicit vs. explicit attitudes

Attitudes are defined as positive or negative feelings toward social objects, such as people, places, and policies (Greenwald & Banaji, 1995). Although traditional expression of prejudice and stereotypes seems to have faded to some extent, especially in progressive westernized societies that promote egalitarian views, research suggests that social prejudice is just more covert and unconscious for an individual, which makes it more difficult to detect and control (Amodio, 2014). Examining such "covert" prejudice thus becomes a methodological challenge. Trying to collect explicit attitudes through surveys may fail, as participants will try to maintain a socially favorable self-presentation due to growing social disapproval of prejudiced beliefs (Hewstone et al., 2002). The issue is especially pronounced for socially sensitive topics (Greenwald et al., 2009). In psychology, this is known as the social desirability bias (Crowne & Marlowe, 1960; Edwards, 1957; Krumpal, 2013). This problem may be exacerbated during in-person experiments where

being observed, or simply just being aware that the experimenter is present outside the experiment booth, may motivate participants to modify their answers, even when explicitly told that 1) the experimenter is not able to see their answers when they are filling out the surveys and 2) their data will be anonymous and that they will not be identifiable in any reports. In addition to that, studies in social psychology have pointed out the severely limited introspective access the participants have into their behavior–i.e. the participants may be genuinely unaware of their biases and how those affect their judgment (Nisbett & Wilson, 1977). Yet, these biases matter–research has convincingly demonstrated that our positive or negative attitudes toward something have implicit effects on processing (Ito & Cacioppo, 2000; Satpute & Lieberman, 2006).

Explicit attitudes have already been shown to affect the processing and interpretation of non-native speech. Fraser & Kelly (2012) found that the subgroup of participants who held negative attitudes toward other ethnic groups rated accented speech as being difficult or impossible to understand and were less accurate in transcribing it. In a similar vein, Ingvalson et al. (2017) showed that listeners' ratings of foreign-accented speakers modulated the accuracy of their speech recognition. Caution is advised when interpreting the findings of these studies causally. As argued by Dragojevic et al. (2017), reduced processing fluency may lead to negative attitudes independently of the content of the message, either directly or indirectly via affect. Thus, one can logically argue that more difficulties with processing a particular accent during one's lifetime may eventually result in more negative affect/attitudes toward that accent, and so a correlation between negative attitudes and processing accuracy becomes the "chicken or egg" paradox. In other words, correlations found in those studies do not shed light on the causal relationship between the two, and more research is needed. In line with that, Cargile et al. (1994) state that "attitudes about language are not a singular, static phenomenon. Rather, they affect, and are affected by, numerous elements in a virtually endless, recursive fashion" (p. 215). Importantly, the above studies only examined the phonetic level of speech processing. However, real-life interactions also require discourse models, pragmatic inferences, mental state attributions, meta-representations, and other complex linguistic and psychological processes.

The above findings provide initial evidence for the importance of explicit attitudes in non-native speech processing, although the problem that self-report measures will not yield honest responses remains. One way to detect covert (implicit) biases is the Implicit Association Test (IAT)

(Greenwald et al., 1998) that does not involve explicit questionnaires. The IAT assesses the ease of association between a bipolar target concept (e.g., native/foreign) and a bipolar attribute (e.g., good/bad). The test is based on two core concepts—an associated semantic network and automatic activation (Hinton, 2017). It is assumed that more associated concepts are connected by stronger neural links than less related ones, which affects the speed and ease of lexical activation (A. M. Collins & Loftus, 1975). This entails that two strongly associated concepts mapped to the same button will yield quicker reaction times (native/good and foreign/bad) than two less associated concepts (native/bad and foreign/good), indicating a cognitive bias (cf. Hinton, 2017 who argues against the idea of a cognitive bias, suggesting that implicit attitudes simply reflect predictive mechanisms of the brain that grew up absorbing cultural associations prevalent in an individual's culture). One study using two versions of the IAT, conceptual and auditory, found that there are indeed measurable biases against non-native speech: participants took longer to pair the concept of a foreign accent with positive than with negative attributes, and that held regardless of the type of accent (strong/weak, attractive/unattractive, etc.) (Roessel et al., 2018). Another study using the IAT in Western Canada showed that participants implicitly preferred Italian to Mandarin Chinese accents, even though explicitly no such preference emerged (MacNair, 2021).

To the best of my knowledge, the IAT measuring the generalized implicit bias against the category of non-native accents (akin to Roessel et al., 2018) has never been used as a predictor of linguistic processing before. Variations of it have, however, been occasionally used in linguistics and other disciplines. For instance, Babel (2010) used the IAT to examine whether implicit Australia and New Zealand biases affected phonetic accommodation. Other studies found that a greater Asian-foreign bias (slower reaction times when Asian faces are paired with American compared to foreign scenes) predicted greater difficulties in transcribing Korean-accented speech in background noise (Yi et al., 2013) and a greater BOLD (blood-oxygen-level-dependent) response in the primary auditory cortex to Korean-accented speech compared to native English speech (Yi et al., 2014) (although only when video recordings of the speakers were presented). Curiously, Romero-Rivas et al. (2022) examined whether implicit biases against Spanish-accented English could predict sentences given to native English and Spanish-accented defendants on trials. While they found that participants were more likely to give significantly harsher sentences to foreign- than native-accented defendants in general, the results of the IAT interacted with those sentences

in a somewhat counterintuitive fashion. The larger the implicit bias *in favour* of a particular accent, the harsher the sentence given to the defendant with the corresponding accent. The authors explained this finding through a "black sheep effect" that describes how people tend to judge their ingroup members much harsher than outgroup members when their behaviour violates expectations. While it is not entirely clear how this would explain the tendency to give harsher sentences to foreign-accented speakers by participants who had an implicit bias in favour of foreign accent (since all participants in the study were native English speakers and thus foreign-accented speakers would generally be considered their outgroup), this finding is still interesting and noteworthy. Of note is also the fact that the IAT test in this study did not use attributes belonging to any particular dimension (e.g., intelligence) but rather positively vs. negatively valenced words ("happy, art, smooth, dream, peace, rose, good, heart" vs. "fear, hate, rage, fury, wrath, fight, hatred, sick"). Finally, as discussed in section 1.2.3, research has found that bilingual individuals show more implicit pro-Moroccan attitudes when they complete the test in Arabic than French, and more pro-Hispanic attitudes when completing the test in Spanish than English (Ogunnaike et al., 2010).

While it has been argued that implicit and explicit measures reflect different aspects of social attitudes and may thus not be particularly correlated (Dovidio & Gaertner, 2010), a meta-analysis based on 126 studies showed that they tend to be (Hofmann et al., 2005). Notably, the authors argued that this correlation strengthens when the two measures share the same underlying construct and when people respond with a higher level of spontaneity in explicit surveys. Research using both of these measures is thus needed to see which of them is more predictive of which aspect of language processing.

There are two options for investigating implicit biases against foreign accents: a conceptual and an auditory IAT. Roessel et al. (2018) showed that the two produce converging results. In the conceptual version, written terms for concepts are used—e.g., AUSTRALIA vs. NEW ZEALAND in Babel (2010) or HOCHDEUTSCH (Standard German) vs. AKZENT (accent) in Roessel et al. (2018). For our purposes, the terms CANADIAN ENGLISH/FOREIGN ACCENT will be used, which should reflect overall prejudice against foreign-accented speakers on the chosen dimension (e.g., knowledgeability or another status variable). Importantly, since no particular accent is used, the results should reflect a non-specific foreign-accent bias (although, of course, this bias could be

shaped by the frequency of a particular accent in each participant's environment and their attitudes toward those accents). Also importantly, both of these terms refer exclusively to language and thus cannot be confused with nationality, which helps avoid potential confounds (K. A. Lane et al., 2007).

Furthermore, importantly for our purpose, it has been argued that the salience of explicit and implicit attitudes depends not only on speaker's characteristics and the situational context, but also on listener's traits, including their social identity, mood, and current goals (Cargile et al., 1994; Cargile & Bradac, 2001). Thus, attitudes represent a complex phenomenon, with both a situational component and a person-based component.

1.3.2 Political orientation

In addition to being their own thing, attitudes are also part of belief systems. The term "belief system" is generally understood as a set of functionally interdependent ideas and attitudes (P. E. Converse, 1964). According to social identity theorists (Tajfel et al., 1971; Tajfel & Turner, 1986), people are naturally drawn to mass belief systems as they foster a sense of belonging. One such system is political orientation along the left-right or liberal-conservative[4] continuum. Jost et al. (2003) and Jost & Amodio (2012) argued that at the heart of the left-right distinction lay two core differences, which have been found correlated: 1) embracing versus rejecting social change, and 2) rejecting versus accepting social inequality. While individuals fall in various places on those scales, how exactly they fall there remains unclear.

Seminal work in political science has convincingly demonstrated that the majority of people do not form their political views ideologically, i.e. through deductive reasoning from a set of integrated social and political principles (P. E. Converse, 1964; Feldman, 1988). To answer the question of how, in that case, such views arise, Jost and colleagues proposed to view political leaning as "motivated social cognition" (Jost et al., 2003, 2007; Jost & Amodio, 2012). In a nutshell, the theory assumes that political views reflect an individual's epistemic, existential, and rational motivation, i.e. how they deal with uncertainty, threat, and conformity. Essentially, sociopolitical attitudes are treated as but one manifestation of a cognitive and affective style and

[4] I will use these terms interchangeably throughout this dissertation.

thus tend to be relatively stable and internally consistent across disparate issues and across lifespan (Block & Block, 2006). This idea has gotten ample empirical support.

To begin with, research shows that conservatives are less tolerant of ambiguity in a broad range of contexts (including low-level visual illusions), have weaker conflict-monitoring activity and thus less neurocognitive sensitivity to conflict (Amodio et al., 2007; Carraro et al., 2014). It is thus reasonable to assume that processing ambiguous language could also be affected by political leaning. In addition, political views have been found to correlate with epistemic needs, so that conservatism is associated with higher personal need for closure, structure, and order (Chirumbolo et al., 2004)—a correlation that strengthens when individuals have better political expertise (Federico & Goren, 2009)—than liberalism. In addition, conservatism is notoriously associated with elevated threat perception and stronger automatic selective attention to negative stimuli (Jost et al., 2003; Oxley et al., 2008). For language processing, this means that high authoritarians[5] (i.e. individuals aligning with right-wing authoritarianism, or RWA; Altemeyer, 1996) but not liberals exhibit faster lexical decision times for threatening words like "cancer" or "poison" than to non-threatening words like "potato" or "telescope" (Lavine et al., 2002). The authors also show that, when compared with liberals, priming with "arms" in high authoritarians provides stronger activation of "weapons" rather than "legs". An argument may thus be made that threatening concepts are more cognitively accessible for conservatives than liberals, resulting in faster lexical access.

Although situation-based components constitute a big portion of intergroup attitudes (Choma & Hodson, 2008), individual differences, including political orientation, have also been shown to play a significant role (Hodson & Dhont, 2015). It has been demonstrated that conservative individuals more readily form illusory correlations between negative information and minority groups—both explicitly (Carraro et al., 2011; Castelli & Carraro, 2011) and implicitly (Carraro et al., 2014). Unsurprisingly, political orientation (and RWA in particular) is also predictive of anti-immigrant attitudes even despite sufficient exposure and immigrant friends (Kiehne & Ayón,

[5] Right-wing authoritarianism is a set of attitudes that revolve around deference to in-group authorities and desire to punish marginalized groups who violate in-group norms, as well as putting conformity above personal autonomy (see Osborne et al., 2023 for a great review of psychological causes and evolutionary antecedents of authoritarianism). While left-wing authoritarianism also exists (Costello et al., 2020), it is much more widespread on the political right.

2016). Authoritarianism has also been linked to higher resistance against non-native speakers in academic settings (Bresnahan et al., 2002).

In summary, political views have been shown to affect a broad range of cognitive and affective processes, including resolution of ambiguity and dealing with uncertainty.

1.3.3 Other character traits

In addition to the traits mentioned above, other traits, both social and cognitive, will be examined. It has been reported in the language attitudes literature that variables like socioeconomic background and bilingualism significantly interact with stereotyped impressions of language varieties and with social evaluations of the speakers of these varieties (Lambert, 1967; Lambert et al., 1960). It would thus be reasonable to hypothesise that variables mediating language attitudes could also mediate linguistic judgments about one's speech. For instance, diversity variables like the self-reported amount of interactions with non-native speakers, general ethnic and linguistic diversity of the participant's childhood and current environment could be collected for that purpose. As one relevant example, Dewaele & McCloskey (2015) showed that childhood exposure to different ethnicities and multilingualism, the diversity of one's workplace, speaking multiple languages, and, surprisingly, being older are correlated with being less bothered by foreign accents.

As to the processing of figurative language and to pragmatic inferences in general, large interpersonal variation has been attested in the literature (Abraham et al., 2021; Olkoniemi et al., 2016; Stamenković et al., 2019; Werkmann Horvat et al., 2022). Although most research has historically focused on the Autism Quotient (AQ) scores and working memory measures, Mayn & Demberg (2022) examined a wider range of individual differences in a pragmatic reference game and found that the Cognitive Reflection Test scores (Frederick, 2005) significantly predicted pragmatic abilities (but not the AQ scores or working memory). For irony, it can be hypothesized that need for cognitive closure defined in clinical psychology as a person's desire to get straightforward, unambiguous answers (Frenkel-Brunswik, 1949) may modulate its detection aptitude. As verbal irony is interpretively ambiguous and as the resolution of ironic ambiguity seems to proceed similarly to other types of ambiguities in different modalities and linguistic aspects (Pexman, 2008), greater need for cognitive closure may be correlated with more taxing processing of ironic statements, especially foreign-accented. For metaphors, individual differences

in creativity, working memory capacity and need for cognition (NFC) (Cacioppo & Petty, 1982), non-verbal intelligence (as measured by the nonverbal Raven's Progressive Matrices test; Arthur & Day, 1994) and multilingualism have been reported (Abraham et al., 2021; Olkoniemi et al., 2016; Stamenković et al., 2019; Werkmann Horvat et al., 2022). Individual differences that may affect resolution of epistemic uncertainty (i.e. facts of uncertain truth) have largely not been investigated before, apart from factors like general accent intelligibility and experience with a particular accent (Jiang et al., 2020).

In summary, a wide array of social and cognitive factors will be investigated as previous literature suggests their involvement in language attitudes, strength of intergroup bias, and pragmatic abilities. Next section discusses methodological challenges this dissertation faces in more detail.

1.4 Methodological considerations

Before I proceed to the research questions of this dissertation, it is important to address the issue of methodology in researching the effects of non-native speaker identity on language processing. The simplest and the most straightforward way is to simply record stimuli spoken by a native and a non-native speaker (as an example, Hanulíková et al., 2012; Romero-Rivas et al., 2015). This method, while being the highest in authenticity, has flaws. First of all, it makes it near-impossible to disentangle the effect of processing difficulty due to reduced intelligibility from expectations of lower LX proficiency and from outgroup categorization. Second, any native and non-native speakers are likely to differ on many phonetic levels, including tone of voice (e.g., voice quality, speaker confidence, etc.), and listeners are notoriously sensitive to such cues (Jiang & Pell, 2015; Phillips et al., 1998; Sauter et al., 2010; Scott, 2019). Specifically, it has been shown that speaker believability ratings are directly affected by perceived speaker confidence (Jiang et al., 2020). To solve this problem, sociolinguistics and social psychology often resort to a matched guise technique (Lambert et al., 1960). In the most relevant variation of this method, the same participant records stimuli in their native accent and imitates a non-native accent. This addresses the concerns expressed above but also suffers from drawbacks, such as unnaturalness that may be consciously and/or subconsciously detected by the listener, or susceptibility to idiosyncrasies of a particular speaker (how well they produce the guise). In addition, a well-performed non-native accent should still induce a higher cognitive load than a native accent, which usually comes with a negative affect

(Dragojevic & Giles, 2016), hence expectations related to lower LX command or social categorization remain confounded with processing fluency.

To address this issue, some studies employ auditory stimuli read by a native speaker but devise a cover story that the stimuli were originally created by a non-native speaker (Foucart & Hartsuiker, 2021). Another method that has recently been gaining popularity is to use exclusively written sentences as stimuli but introduce sentence "authors" as native or non-native speakers through a written vignette presented either before the experiment or before each block of stimuli (Bazzi et al., 2022; Fairchild et al., 2020; Foucart et al., 2019; Lorenzoni et al., 2022). The sentences are then presented together with the picture or simply with a bio-description of their purported "author". These methods completely remove processing disfluency from the equation but, naturally, have their drawbacks too. Both methods are much riskier in terms of authenticity and believability. Although extra visual cues may be used, such as presenting a country's flag to indicate the speaker's origin (Lorenzoni et al., 2022) or an ethnic name (Fairchild & Papafragou, 2018), it is still much easier to ignore this information in favor of focusing only on the written sentence. A blocked design may be beneficial as it allows to not divide the reader's attention between multiple people at the same time but may not work for methods like EEG.

All in all, none of the methods provide a comprehensive picture. Ideally, auditory presentation of stimuli should be compared to written delivery of the same stimuli to shed light on how much of the effect comes from speech disfluency and how much from expectations triggered by the non-native speaker identity. Thus, this dissertation uses both methods to investigate the resolution of linguistic as well as epistemic uncertainty.

1.5 The present dissertation

The present dissertation aims to contribute to and expand on previous research by examining the effects of non-native speaker identity on comprehension of figurative language and language with unknown truth values, in other words, language associated with interpretative uncertainty. The overarching goal is twofold. The first goal (1) is to examine whether incorporating the identity of a non-native speaker has pragmatic implications manifested in offline linguistic judgments, i.e. speaker effects. Specifically, I want to examine whether pragmatic inferences from figurative

language and truth evaluation of unknown facts are hindered for non-native compared to native speakers. The second goal (2) is to examine whether these processes are affected by individual differences, i.e. listener effects. As discussed above, the major focus of this thesis is on traits that shape a person's response to the outgroup and their place in the social hierarchy, such as explicit and implicit attitudes toward non-native speakers, political and social dominance orientation, and empathy. In addition, because of interpretative ambiguity of the chosen stimuli, I aim to examine several cognitive (need for cognitive closure and cognitive reflection) and personality (Openness and Extraversion) traits. An important part of this research agenda is to test whether explicit accent attitudes are correlated with implicit accent attitudes and whether either of them (or both) affects the pragmatic interpretation of non-native speech. Generalized implicit accent prejudice has never been investigated as a psycholinguistic predictor before; hence this dissertation aims to fill this knowledge gap.

Answering the research questions stated above will help bridge disciplines such as sociolinguistics (i.e. language attitudes), social psychology (i.e. social cognition, including social evaluations of the speaker), linguistics (i.e. linguistic manipulations of the stimuli and (meta)linguistic judgments about those stimuli), and personality psychology (i.e. individual differences in character traits). The findings from this research can, in turn, be used to investigate subsequent *social* inferences about the speakers—inferences based on how exactly the speaker used language in a specific interaction, which produce a cluster of evaluations regarding the speaker's identity, demographics, personality and ideology (Beltrama, 2020; Beltrama et al., 2023; Beltrama & Papafragou, 2023; Eckert & Labov, 2017). In addition, successful intersection of these disciplines will contribute to taking interpersonal differences in language processing into account rather than treating them as error variance, and will help refine language comprehension models that assume meaningful speaker and listener effects (Münster & Knoeferle, 2018; van Berkum, 2018). Finally, the findings of this dissertation may serve as a stepping stone to investigating speaker and listener effects during online processing of figurative language and facts of uncertain truth using time-sensitive methods such as EEG, fMRI, or eye-tracking/pupillometry.

Chapter 2 (Experiment 1) investigates the perception of non-native irony by native speakers of English. Although the subject remains virtually unresearched, some previous studies (Caffarra et al., 2019; Caffarra & Martin, 2019) suggested that pragmatic inferences from ironic praise, a rare

form of irony, may be hindered when the speaker is non-native. I use dialogues between native Canadian English speakers and their non-native (Mandarin Chinese) peers with more naturalistic prosodic cues to investigate whether the results of previous studies were a by-product of unnaturally slow speech rate and flat intonation, or an effect driven by the speaker's linguistic identity. In addition, I examine several other rating scales in addition to irony rating (listener's confidence in their interpretation, remark appropriateness and offensiveness) to shed more light on how ironic language of non-native speakers is perceived. Crucially, the study also investigates individual differences in these pragmatic judgments (political leaning, ambiguity intolerance, and mentalizing abilities).

Chapter 3 reports a follow-up study on Chapter 2. Specifically, it investigates the prosodic side of non-native irony. To date, no data exists about prosodic contrasts in non-native irony for any accent. Since Chapter 2 reports significant pragmatic implications for both types of non-native irony, it is important to investigate its acoustic correlates. Even when prosodic parameters calculated over entire utterances do not differ between various types of native and non-native irony, a fine-grained analysis with attention to segment-level and word-level parameters may reveal hidden differences. I hypothesise that non-native irony would be more prosodically marked than native irony and that ironic praise would be more marked than ironic criticism, both in order to minimize the risk of misunderstanding.

Chapter 4 (Experiment 2) examines a different type of non-literal language—metaphors. Although this topic has not been investigated before, it can be hypothesised that semantic inaccuracy commonly associated with non-nativeness will affect the perceived sensibility of metaphorical utterances. Importantly, since Chapter 3 found fine-grained prosodic differences between native and non-native speech, I use a fully written modality to eliminate the possibility that processing disfluency and segmental/suprasegmental deviations of non-native speech affect the results. Again, I investigate the effect of interpersonal variation in linguistic judgments: implicit and explicit accent attitudes, political orientation, Extraversion and Openness, and cognitive reflection.

Chapter 5 (Experiments 3 and 4) investigates truth evaluations of world knowledge facts. While a substantial amount of research on this topic exists, previous findings are extremely inconsistent, in no small part due to methodological, conceptual, and linguistic differences between the studies. I aim to contribute to this research by focusing not on generalized prejudice against non-native

speakers across all knowledgeability domains but instead on a specific subset of world knowledge, namely facts related to "ingroup knowledge", such as geographical, political, and cultural facts about Canada. The primary goal of the two experiments reported in this chapter is to investigate whether social categorization of the speaker affects truth ratings over and above the effects of speech disfluency by 1) comparing auditory to written stimuli and 2) manipulating the non-native speaker's group membership by introducing them as either Canadian citizens or recent immigrants. Importantly, I investigate whether there is meaningful interpersonal variation in these judgments by examining the strength of an individual's implicit attitudes toward non-native speakers' knowledgeability and using it as a predictor for ratings.

Chapter 2

I was being sarcastic!: The effect of foreign accent and political ideology on irony (mis)understanding

Abstract

Misunderstood ironic intents may injure the conversation and impede connecting with others. Prior research suggests that ironic compliments, a rarer type of irony, are considered less ironic when spoken with a foreign accent. Using more ecologically-valid stimuli with natural prosodic cues, we found that this effect also applied to ironic criticisms, not just to ironic compliments. English native speakers (N = 96) listened to dialogs between Canadian English speakers and their foreign-accented peers, rating targets on multiple scales (irony, certainty in the speaker's intent, appropriateness, and offensiveness). Generalized additive mixed modelling showed that 1) ironic comments were rated lower for irony when foreign-accented, whereas literal comments were unaffected by accent; 2) the listener's political orientation, but not empathy or need for cognitive closure, modulated irony detection accuracy. The results are discussed in terms of linguistic expectations, social distance, cultural stereotypes, and personality differences.

Keywords: Irony, Foreign accent, Dialog, Political ideology, Individual differences

2.1 Introduction

A foreign accent is an integral and mostly permanent part of a person's social identity, along with such markers as race, ethnicity, or age. Unfortunately, it is these permanent markers that often trigger implicit and explicit biases. Mounting evidence suggests that non-native speakers consistently face negative perception, being judged as less reliable, less credible, less intelligent, and less successful than native speakers (Foucart et al., 2019; Fraser & Kelly, 2012; Fuertes et al., 2012; Lev-Ari & Keysar, 2010). This bias starts in early childhood, with children preferring native over non-native speakers as their friends, and continues into adulthood (Imuta & Spence, 2020; Kinzler et al., 2009). A recent study provided evidence that any foreign-accented speech

immediately and automatically triggers negativity biases simply by virtue of belonging to a non-native accent category (Roessel et al., 2018). From a linguistic point of view, a foreign accent alters multiple levels of language processing, including lexical access, semantic integration, reanalysis, and processing depth (Lev-Ari, 2015; Porretta et al., 2016, 2020).

Since psycholinguistic investigations into foreign accents are recent, virtually all existing research is based on literal language. However, making inferences from non-literal language is also hard, if not harder than from literal speech. Irony, for instance, requires complex pragmatic inferencing, meta-representations, as well as first- and second-order mental state attributions to be properly understood (Colston & Gibbs, 2002; Dennis et al., 2001). Considering what is already known about the processing of foreign accents (e.g., Grey & van Hell, 2017; Hanulíková et al., 2012; Romero-Rivas et al., 2015) and the error-prone nature of foreign-accented speech, we can hypothesise that pragmatic inferencing from ironic utterances may be carried out quantitatively and/or qualitatively differently for non-native speech. As but one example, listeners have lower expectations regarding non-native speakers' linguistic competence (Lev-Ari, 2015), consequently treating their speech as less reliable. When encountering an utterance that is counterfactual to the preceding context (as is the case with most ironies), the listeners may thus have more reason to consider alternative explanations for this discrepancy, for instance that a wrong lexical item was chosen or that a non-native speaker simply misunderstood what happened.

In addition, because of the complexity of mental operations needed to understand irony, there is stable meaningful variance in irony performance among healthy adults, with some detecting ironic comments with almost perfect accuracy and some barely above the chance level (Bruntsch et al., 2016; Winner et al., 1988). Only very few predictors have been identified so far, including schizotypal and borderline traits (Kieckhäfer et al., 2019), trait anxiety (Gucman, 2016), trait bad mood and benevolent humor (Bruntsch & Ruch, 2017). Moreover, the perception of foreign-accented speech is also notoriously affected by individual differences, both at the level of speech recognition (McLaughlin et al., 2018) and affective response (Bresnahan et al., 2002). It thus seems likely that not only will pragmatic inferencing from foreign-accented irony be more difficult than from native irony, but also that it will be subject to significant interpersonal variation in both cognitive and affective traits.

To the best of our knowledge, the only two studies examining the perception of foreign-accented irony were conducted on native and English-accented Spanish (Caffarra et al., 2018, 2019). In the rating study (Caffarra et al., 2018), ironic praise—but not criticism —was considered less ironic when spoken with a foreign accent. The authors speculated that foreign-accented ironic praise was more often taken at face value since it is an infrequent trope that non-native speakers are not expected to master. The ERP study (Caffarra et al., 2019) found a larger N400 effect for native ironic praise only and a longer-lasting P600 effect for ironic utterances, which interacted with the type of irony and marginally with the accent. The researchers attributed the lack of an N400 effect for foreign-accented irony to the lack of anticipatory processes (cf. Porretta et al., 2020; Romero-Rivas et al., 2016). These findings are intriguing and in line with the mounting literature showing that listeners actively use information about the speaker to interpret their speech (Kamide, 2012; Kroczek & Gunter, 2021; Lev-Ari, 2015; Romero-Rivas et al., 2015; Tesink et al., 2009; van den Brink et al., 2012a). It will, however, be useful to also consider potential sources of this effect in experimental design choices. For instance, the speakers in those experiments were instructed to maintain a flat intonation and slow speech rate for both literal and ironic stories. Native speakers were first asked to listen to the stories produced by non-native speakers to mimic their speech rate. This is understandable from a standpoint of experimental control but makes speech in general and ironic comments in particular less natural. As languages, including Spanish, commonly use tone of voice and intonation to convey irony (Escandell-Vidal & Prieto, 2020), unnatural use of prosody might make the results less informative of and less generalizable to everyday language use. In addition, the participants were tasked with rating every story first for accent strength and intelligibility, and only then for irony. Such procedure is likely to draw attention to the foreign accents, making them very salient. Rating foreign accents for their strength may also prompt the use of meta linguistic judgements in subsequent irony ratings, blurring the initial gut response.

Assuming that the effect persists regardless of experimental design, one may then ask what the possible reasons for non-native ironic compliments to be rated as less ironic than native ones are. Ironic praise does not require any linguistic skills over and above what is required for ironic criticism, and in most cases the two types are structurally identical (consider: "*It is known that he cooks very well/badly*", Caffarra et al., 2018, p. 4). To date, we are not aware of any studies examining whether ironic compliments employ different prosody than ironic criticisms, although

such research would certainly be warranted. In light of the above, it becomes unlikely that foreign-accented speakers cannot, or would not be expected to, master ironic praise. This is very different from, say, idioms or grammatical agreement, which indeed require special knowledge and are language-specific (Cieślicka, 2015; Hanulíková et al., 2012). Further, irony seems to be a universal phenomenon, and it is its prosodic marking that differs cross-linguistically (González Fuente, 2017). If this is indeed the case, one would expect this pragmatic knowledge to be transferable to a new language.

That said, it is nevertheless possible that non-native speakers are less likely to use ironic praise for at least two reasons, one having to do with "the asymmetry of affect" (Clark & Gerrig, 1984) and the other with the social circumstances in which it is most commonly used. "The asymmetry of affect" refers to the fact that ironic compliments, even though used in the face of positive circumstances, evoke more negative feelings than ironic criticisms. The reason for that asymmetry stems from the surface form of ironic compliments, which violates conversation etiquette and politeness expectations ("*You are such a rotten friend!*" said to someone who has been nothing but loyal to you). In Dews, Kaplan, et al. (1995), participants rated ironic praise as more insulting than its literal counterpart, whereas ironic criticisms were rated as less insulting than literal ones. Since foreign-accented speakers are usually aware of and expect stigmatization by others due to their accent (Derwing, 2003; Gluszek & Dovidio, 2010), they might avoid language that can be perceived as rude and exacerbate this negative perception. The risk is much higher than in the case of ironic criticisms, since the latter could just be taken as literal compliments if misunderstood. Second, several studies demonstrated that common ground and closeness of the speaker–addressee relationship is much more important for the proper interpretation of ironic compliments than they are for ironic criticisms (Bruntsch & Ruch, 2017; Pexman & Zvaigzne, 2004). In addition, Kreuz & Link (2002) showed that ironic praise is much easier to interpret with an explicit antecedent event, whereas ironic criticism can be based on purely implicit antecedents (such as unstated societal/cultural norms or expectations). What that means is that one can use ironic criticism ("*You are indeed a great chef!*") without explicit reminders of what was expected (a good meal). In contrast, ironic praise ("*You are indeed a terrible chef!*") without an explicit antecedent (for example, someone previously complaining that they cook terribly and then preparing a delicious meal) is harder to interpret. All this suggests that the speaker and the addressee need to be close

enough and have enough shared background to be "entitled" to use ironic compliments without the risk of injuring the conversation. Non-native speakers may simply not feel close enough with native speakers to use it and/or have fewer antecedents to refer to. According to the "heuristic of inferability" proposed by Kreuz (2018), an individual's likelihood of using irony depends on how certain they are that their irony will be correctly understood. If nonnative speakers are not really certain that their ironic compliments will not be taken at face value because their relationship with the addressee is not close enough, they may be less motivated to use them. Listeners, in their turn, may automatically assume less common ground and thus more social distance between native and non-native speakers (unless explicitly stated or demonstrated otherwise). This assumption of a larger social distance may, consciously or subconsciously, make native listeners consider foreign-accented ironic compliments less appropriate. Summarizing, it seems crucial for ironic praise to be grounded in sufficient context and be used between people who are close enough in order to be understood.

We attempted to address the limitations of prior research by using more ecologically-valid stimuli—dialogs between native and foreign speakers conversing as peers equal in social status. Since the primary function of ironic compliments is to be playful and to tease, and since the speaker-addressee relationship appears to be very important in this case, we constructed the dialogs in such a way that their style suggested that our speakers were either close acquaintances or friends. This should facilitate metarepresentational inferences for the listeners, as they can assume that the speakers have enough common ground to actually use irony without injuring the conversation. Additionally, our speakers used natural prosody to further facilitate irony interpretation. Native speakers of English have been shown to rely on prosodic cues to identify sarcasm, of which the primary ones are slower speech rate, greater intensity, and lower pitch level (Rockwell, 2000). It has also been shown that adult listeners are able to correctly identify irony based on prosody alone (with no context provided), further testifying to its importance (Mauchand et al., 2019; Rockwell, 2000). To examine whether the type of irony has a bearing on listeners' ratings, we also used criticism (sarcasm) and praise (teasing).

We had two hypotheses for this study. First, based on the reasons discussed above, we predicted that foreign-accented irony of any type will be rated less ironic when spoken with a foreign accent even when sufficient contextual and prosodic cues are available for disambiguation. Since ironic

compliments are harder to interpret even for native speech, they should be proportionally more difficult when delivered in a foreign accent. Second, we expected individual differences in the need for cognitive closure, empathy, and political ideology to modulate irony detection aptitude. Motivation and more specific predictions for each character trait are outlined below.

Since irony comprehension requires both cognitive and affective processes, it should be modulated by traits with cognitive and affective components. Additionally, given that foreign-accented irony is produced by non-native speakers, traits that correlate with anti-immigrant biases may shape its detection and processing. Political attitudes have been shown to be a robust predictor of anti-immigrant prejudices (Banton et al., 2020; Hodson & Dhont, 2015). Aligning with political conservatism is predictive of anti-immigrant attitudes even despite sufficient exposure and immigrant friends (Kiehne & Ayón, 2016). We thus hypothesized that more conservative listeners may invest less effort in understanding foreign-accented speakers and thus miss their irony more often than less conservative listeners. We measured participants' political ideology along the left-right dimension using the 20-item Wilson-Patterson Conservatism Scale (W-P) (Wilson & Patterson, 1968). The need for cognitive closure, which clinical psychology defines as a person's desire to get straightforward, unambiguous answers (Frenkel-Brunswik, 1949), may also modulate irony detection aptitude. Verbal irony is inherently ambiguous since the true meaning of an ironic utterance is often the opposite of its surface meaning, but the listener needs to decide which meaning was actually intended. There is accumulating evidence that resolution of ironic ambiguity might be akin to other types of ambiguities, for instance in visual imagery, word recognition, or syntactic parsing (Pexman, 2008), when multiple cues seem to be assessed and weighted in parallel. Naturally, the listener needs to first recognize that the intended meaning of an utterance might be different from its literal meaning, and the clash between the preceding context and the speaker's comment (*"You are really a terrible driver!"* said after you passed your driving exam with flying colors) usually provides an initial indication. Given this ambiguity, greater need for cognitive closure may be correlated with more taxing processing of ironic statements, especially foreign-accented. Individuals with a greater need for cognitive closure also tend to attain it as quickly as possible and maintain it for as long as possible, which may manifest in them being more confident in their interpretation. This trait was measured using the 47-item Need for Cognitive Closure scale (NFCS) (Webster & Kruglanski, 1994). Lastly, irony requires the ability to mentalize

other people's states of mind. A lesion study by Shamay-Tsoory et al. (2005) has shown that patients with impaired empathy are also impaired in verbal irony comprehension, but it is not clear whether empathy effects extend to non-clinical populations. We predicted that participants with lower empathy might take ironic statements as literal more often than those with higher empathy, and even more so for foreign-accented speech. There is also evidence that empathy correlates with the "pragmatic N400" indexing the participant's sensitivity to violations of social stereotypes (van den Brink et al., 2012a). Importantly, this study also showed that participants with higher empathy adapt faster to pragmatic stimuli and adjust their expectations for particular content once they have sufficient evidence that social stereotypes are no longer a reliable cue. Thus, participants with higher empathy may show stronger learning effects for foreign-accented irony during the experiment, resulting in an even more accurate performance as the study progresses. As a proxy for this trait, we used the Interpersonal Reactivity Index (IRI) developed by (Davis, 1980). Overall, since ironic praise violates politeness expectations in its surface form and is more cognitively demanding (Bruntsch et al., 2016; Hancock et al., 2000; Harris & Pexman, 2003), we expected its comprehension to be more shaped by the listeners' individual differences.

2.2 Materials and methods

2.2.1 Participants

Ninety-six self-identified native English speakers participated in the experiment. All participants were students and received partial course credit for their participation. Data from 3 participants were excluded due to failed attention checks (N = 2) or an unacceptable lie score in the NFCS survey (N = 1). The final sample included 93 participants (11M [12%], mean age: 20.4, range: 17–48, SD: 4.7). The mean self-reported English proficiency was 4.7 out of 5.

2.2.2 Materials

Thirty sets of dialogs were constructed as experimental items and fifteen more as fillers. Since all experimental targets directly commented on the speaker's actions and had similar structure, with an evaluative item being in the middle ("You are indeed wasteful with your finances" or "You always were incredibly mature in your behavior"), filler items contained neutral remarks with

random sentence structure ("I'm glad environmental issues attract so much attention", "You can get a new one at Home Depot at a bargain price", or "Money is tight at the moment so I appreciate the suggestion") to hide the purpose of the experiment. Each set of experimental dialogs consisted of eight conditions, which yielded a crossed 2 × 2 × 2 design (Table 2.1). Since irony perception is notoriously affected by how strong the contrast between the preceding context and the ironic remark is, all our dialogs built strong expectations for a particular reaction that was subsequently violated (for example, someone describing that they were literally kicked out of their cello class for being "the worst cello player ever" rather than saying that they just accidentally played out of key). No conventionalized ironic remarks were used ("Very funny"), and in this respect all our ironic utterances were "novel" (see Giora, 1999 for the idea that the salient, lexicalized meaning of conventionalized ironies is the ironic one, which affects their processing). Most dialogs contained intensifying lexical markers such as superlative adjectives and exaggeration adverbs (for example, "the most", "really", "indeed", "surely") that are commonly used in ironic remarks and bias listeners for non-literal interpretation (Ackerman, 1983). Eight experimental lists were created by simple item rotation to ensure that the participants only heard each item in one condition. Due to technical issues with Pavlovia at pavlovia.org, lists had to be pseudorandomized, and every participant assigned to the same list heard items in the same order. All lists started with two fillers.

The dialogs were recorded directly onto a computer hard disk, with the speakers sitting in adjacent sound-treated WhisperRoom booths using Countryman Earset microphones. The audio was saved in the WAV format with a 44.1 kHz sampling rate and a 16-bit precision. Two native speakers of Canadian English, from Ontario and Alberta, and two native speakers of Mandarin Chinese from Mainland China read the dialogs out loud. All speakers were male and in their mid-twenties. All were students at the University of Alberta. Each native speaker was paired with both foreign-accented speakers so that each pair of speakers read a quarter of the stimuli. Unlike Caffarra et al. (2018), we opted for natural prosody and let the speakers read ironic utterances with intonation they would normally use for conveying irony. The percentage of Canadians who identify themselves as being of Chinese origin is high in Edmonton (estimated at around 7% according to Statistics Canada, 2017), which means that our participants are likely to have had sufficient exposure to Chinese-accented English and should be relatively familiar with its prosody. The order

in which the participants started reading the dialogs was counterbalanced to avoid the speakers mimicking each other.

Table 2.1. The list of experimental conditions with examples. Every dialog in this table was recorded twice, with the native and foreign-accented speakers swapping roles. The order in which the speakers started reading the dialogs was counterbalanced. Target words are in bold.

Irony Type	Ironic	Literal
Praise	—How did the dinner with your wife's parents go? —I spent the whole day cooking salmon and making drinks. They finished everything, down to the last crumb, and asked for seconds. —You surely are the most **terrible** cook in this part of town	—How did the dinner with your wife's parents go? —I spent the whole day cooking salmon and making drinks. They finished everything, down to the last crumb, and asked for seconds. —You surely are the most **talented** cook in this part of town
Criticism	—How was your Saturday dinner with your wife? —I decided to try a new fish recipe. I probably messed up because she couldn't finish it and the next day got really bad food poisoning. —You surely are the most **talented** cook in this part of town	—How was your Saturday dinner with your wife? —I decided to try a new fish recipe. I probably messed up because she couldn't finish it and the next day got really bad food poisoning. —You surely are the most **terrible** cook in this part of town

After the recording, two sets of dialogs were excluded because the target word was pronounced incorrectly. Seven English speakers were then asked to transcribe the foreign-accented target utterances taken out of context. Since all our items were spoken twice, with a literal and an ironic intonation, listeners were divided in two groups so that none of them would hear the same word twice. Four more sets were excluded because the mean agreement for the same target word was 50% or less. This left us with 24 dialog sets (192 dialogs; all the stimuli are available in Appendix A.1).

2.2.3 Acoustic analysis

We first scaled the intensity of the dialogs to 70 dB. We then calculated mean duration, mean pitch, mean speech rate (the number of syllables per time in seconds; de Jong & Wempe, 2009)

and pitch range of the target utterances (Table 2.2) using PRAAT (6.11.16) (Boersma & Van Heuven, 2001).

Table 2.2. The comparison of acoustic features between the conditions. SDs are reported in parentheses.

Condition			Prosodic parameters			
Irony	**Type**	**Accent**	**Mean duration, ms**	**Mean pitch, Hz**	**Mean speech rate, nsyll/sec**	**Pitch range, Hz**
ironic	criticism	native	2459 (305)	129 (8)	4.20 (.54)	112-144
		foreign	2639 (378)	113 (12)	4.17 (.67)	91-131
	praise	native	2523 (323)	126 (6)	4.32 (.59)	112-143
		foreign	2871 (460)	116 (17)	4.15 (.69)	92-145
literal	criticism	native	2247 (282)	133 (16)	4.41 (.83)	113-183
		foreign	2435 (363)	110 (14)	4.32 (.71)	91-135
	praise	native	2194 (282)	130 (7)	4.49 (.78)	114-145
		foreign	2506 (393)	111 (15)	4.28 (.59)	90-137

For each dependent variable (duration, pitch, and speech rate) we fitted three increasingly complex mixed-effect models using the *nlme* package (3.1.149) (Pinheiro et al., 2021) in R (4.0.3) (R Core Team, 2020). The data with all the scripts are available on OSF. The first model included one random effect of item. The next model added irony, type, and accent as predictors. A three-way interaction was added last. For both duration and pitch, the three predictors significantly improved the model's fit, whereas the interaction decreased it. The results of the final models are summarized in Table 2.3. We omit the speech rate model because no predictor was significant. Consistent with previous studies (e.g., Caillies et al., 2019; Regel et al., 2011), ironic utterances were longer in duration than literal utterances. Importantly, foreign-accented ironic comments were longer in duration than their respective literal comments (2639 ms vs. 2435 ms for criticism and 2871 ms vs. 2506 ms for praise). Further, foreign-accented utterances were longer and lower-pitched than native ones. Pitch between ironic and literal utterances did not differ. There were no significant interactions between irony and accent.

2.2.4 Procedure

The experiment was programmed in PsychoPy3 and run on Pavlovia. The participants were asked to listen to the dialogs one by one and then rate the last utterance of each dialog. The rating questions were as follows:

1. How appropriate is the speaker's utterance?
2. Is the speaker being ironic?
3. Is the speaker being offensive?
4. How certain are you that you correctly interpreted the speaker's intent?

The scale ranged from 1 "not at all " to 7 "extremely". The participants were explicitly asked to provide "snap" judgements using their gut feelings. After the participant had rated the dialog, the next one started playing automatically. Each participant rated 39 dialogs, of which 15 were fillers. There were two opportunities for taking a break. The individual difference questionnaires were presented after the main experiment. All questionnaires together with their scoring systems and response scales are available in Appendices B.1-B.3.

Table 2.3. Results summary of the best-fitting linear mixed-effects regression models with duration (in ms) and pitch (in Hz) as dependent variables. Asterisks indicate statistical significance (*p < 0.05, **p < 0.01 and ***p < 0.001).

Duration						
	Value	**Std. Error**	**dF**	***t*-value**	***p*-value**	
(intercept)	2712.24	50.96	188	53.22	<0.001	***
irony literal	-277.44	50.96	188	-5.44	<0.001	***
type praise	78.34	50.96	188	1.54	0.126	
accent native	-256.87	50.96	188	-5.04	<0.001	***
Pitch						
(intercept)	112.79	1.83	188	61.61	<0.001	***
irony literal	-0.01	1.83	188	-0.00	0.997	
type praise	-0.39	1.83	188	-0.21	0.833	
accent native	16.69	1.83	188	9.12	<0.001	***

2.2.5 Data analysis

The dataset for this study can be found in [dataset] (Puhacheuskaya & Järvikivi, 2021). All analyses were done in R (4.0.3) (R Core Team, 2020) and are available on OSF. We used generalized additive mixed modelling for ordinal data (GAMM; Wood, 2017), a non-parametric equivalent of a regression analysis which allows one to analyze random effects and individual differences without assuming linearity (Divjak & Baayen, 2017). The only way to examine the effects of individual differences (continuous predictors) together with factorial predictors in a GAMM model is to represent the interaction between factorial predictors as a grouping predictor

with *n* levels (see van Rij et al., 2020). Since accent, irony and type were all discrete predictors, the interaction was modelled as an eight-level grouping predictor Condition (*ironic/literal criticism/praise foreign/native accent*). All models included random smooths for participant and dialog and factor smooths for participant by trial. For each rating type, we did stepwise forward model elimination using the *mgcv* package (1.8.33) (Wood, 2017) and the *compareML()* function from the *itsadug* package (2.4) (van Rij et al., 2020). The *compareML()* function outputs a chi-square test of REML scores and an AIC difference between two models. If the chi-square test had a *p*-value of > .05, suggesting a non-significant difference in REML scores between a less and a more complex model, then the simpler model was preferred, and the predictor was removed. If the predictor was significant in the model's output but the *compareML()* function showed no significant difference in REML scores, we took a conservative approach and removed the predictor. All final models included in the paper showed a significant improvement ($p \leq .01$) over the models with *n*-1 predictors. There were no "borderline cases" with *p*-values in-between .05 and .09. The main reference level for the intercept was ironic criticism native. Since the difference between the literal and ironic conditions was large, we additionally tested whether the difference between accents was significant for each condition (*ironic/literal praise/criticism*) by choosing different native conditions to estimate the intercept. The results were plotted using the *mgcviz* (0.1.6) (Fasiolo et al., 2023) and *itsadug* (2.4) packages. The scores in the individual difference questionnaires were standardized. The correlations (obtained using the rcorr() function from the Hmisc package) were as follows: $r = 0.25$ between W-P and IRI scores ($p = .015$), $r = 0.32$ between W-P and NFCS scores ($p = .002$), and $r = 0.19$ between IRI and NFCS scores ($p = .075$). Cronbach's alphas (obtained using the *ltm* package) were as follows: 0.83 for NFCS, 0.85 for IRI, and 0.79 for W-P. Please note that none of the final models contained more than one individual difference, thus a problem of multicollinearity did not arise.

2.2.6 Results

We started with checking the correlations between the four rating scales. Irony and Certainty were the least correlated ($r = 0.05$, $p = .017$). The rest of the scales were correlated with $p < .001$, with the strongest correlation found between Offensiveness and Appropriateness ($r = 0.58$). Table 2.4 provides the mean ratings for every condition.

Table 2.4. The comparison of mean ratings between conditions. SDs are reported in parentheses.

Condition			Rating Type			
Irony	**Type**	**Accent**	**Irony**	**Appropriateness**	**Offensiveness**	**Certainty**
ironic	criticism	native	6.14 (1.47)	4.56 (1.68)	4.01 (1.73)	5.71 (1.21)
		foreign	5.64 (1.79)	4.62 (1.57)	3.84 (1.63)	5.35 (1.44)
	praise	native	5.25 (2.05)	3.52 (1.95)	4.06 (1.99)	5.34 (1.47)
		foreign	4.75 (2.12)	3.76 (1.99)	3.73 (1.99)	5.02 (1.72)
literal	criticism	native	2.84 (1.92)	4.22 (1.79)	4.73 (1.68)	5.45 (1.31)
		foreign	2.87 (1.94)	4.44 (1.73)	4.28 (1.75)	5.27 (1.51)
	praise	native	2.20 (1.68)	5.99 (1.25)	1.89 (1.27)	5.74 (1.35)
		foreign	2.17 (1.58)	5.81 (1.37)	1.91 (1.30)	5.69 (1.47)

Irony rating. Every condition was significantly different from the chosen baseline in the final model (Table 2.5). As noted above, we additionally tested the difference between all foreign-accented conditions in relation to native ones by re-leveling the model. As predicted, foreign-accented irony was considered significantly less ironic than native one, and that held for both criticism (p = .008) and praise (p = .044). Crucially, accent had no effect on literal conditions (all ps > .7). This indicates that making pragmatic inferences from foreign-accented speech is heavily context-dependent: statements that do not clash with the preceding context are easily classified as literal regardless of accent, while counterfactual statements evoke more uncertainty as to their proper interpretation when delivered in a non-native accent (Fig. 2.1A). Also, consistent with the previous research, ironic praise in general was deemed significantly less ironic than criticism. As discussed in the Introduction, this may be explained by its surface form that violates social norms.

Of the individual difference measures, only political ideology significantly improved the model's fit. Smooth plots in Fig. 2.2 show that right-leaning participants (higher W-P scores) were worse at detecting irony compared to their left-leaning peers: they consistently rated ironic comments as less ironic and literal comments as more ironic. Further, difference plots in Fig. 2.3 show that the results discussed above were significant across the entire spectrum of W-P scores. Importantly, however, our prediction that right-leaning individuals would miss ironic intent in foreign-accented speech more often than left-leaning ones did not prove true. There was also no interaction between the type of irony (praise or criticism) and political orientation.

To make sure the difference in irony rating was not due to intelligibility we also added the results of the intelligibility pre-test as a predictor to the model with Condition (for native speech, we took

it to be 100% for all items).The predictor turned out to be insignificant, both by itself ($\chi 2 = 1.70$, $p = .192$) and in interaction with Condition (all ps > .4).

Table 2.5. Summary of the best-fitting GAMM with Irony Rating as a dependent variable. The model's formula: IronyRating ~cond + s(W-P, by = cond, k = 3, bs = "tp") + s(participant, bs = "re") + s(dialog, bs = "re") + s(trial, participant, bs = "fs", m = 1), family = ocat(R = 7). Deviance explained = 28.1%.

Parametric coefficients	**Estimate**	**Std. Error**	**z-value**	**Pr(>| *z* |)**	
(Intercept) (ironic criticism native)	3.31	0.21	16.13	<0.001	***
ironic criticism foreign	-0.71	0.27	-2.63	0.008	**
ironic praise native	-1.14	0.27	-4.20	<0.001	***
ironic praise foreign	-1.67	0.27	-6.18	<0.001	***
literal criticism native	-3.56	0.27	-13.25	<0.001	***
literal criticism foreign	-3.55	0.27	-13.20	<0.001	***
literal praise native	-4.46	0.27	-16.25	<0.001	***
literal praise foreign	-4.37	0.27	-16.05	<0.001	***
Smooth terms	**edf**	**Ref.df**	**χ2**	**p-value**	
s(W-P): ironic criticism native	1.59	1.81	10.64	0.015	*
s(W-P): ironic criticism foreign	1.25	1.42	10.65	0.006	**
s(W-P): ironic praise native	1.00	1.00	7.70	0.006	**
s(W-P): ironic praise foreign	1.87	1.97	13.73	<0.001	***
s(W-P): literal criticism native	1.00	1.00	5.01	0.025	*
s(W-P): literal criticism foreign	1.57	1.79	2.03	0.422	
s(W-P): literal praise native	1.80	1.95	17.70	<0.001	***
s(W-P): literal praise foreign	1.00	1.00	13.25	<0.001	***
s(participant)	28.98	91.00	41.56	<0.001	***
s(dialog)	114.33	184.00	300.68	<0.001	***
s(trial,participant)	63.31	835.00	153.97	0.006	**

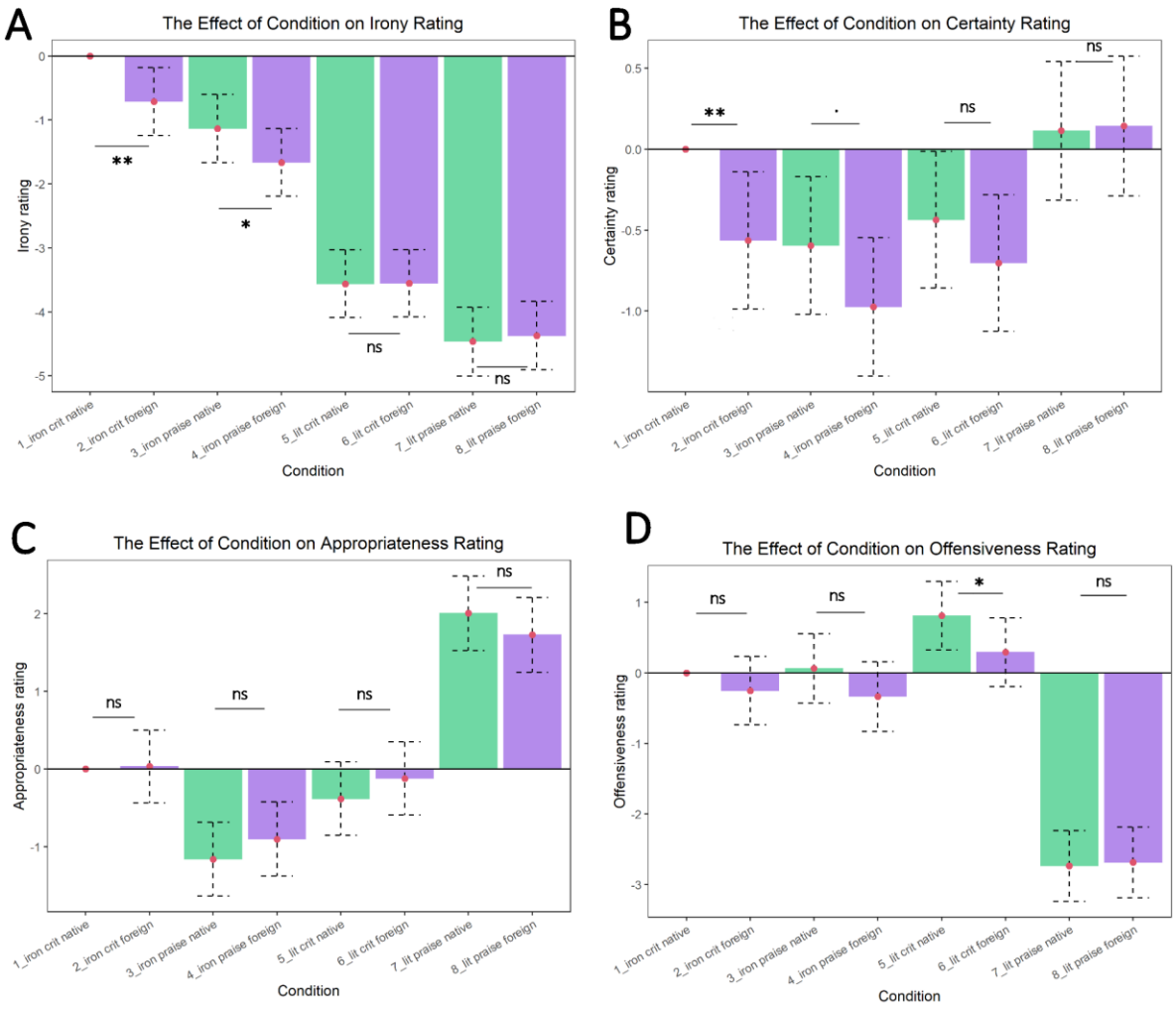

Figure 2.1. The parametric effects plots for all rating types (irony, certainty in the speaker's intent, appropriateness, and offensiveness). Stars indicate significant differences against the intercept (< 0.001 "***", < 0.01 "**", < 0.05 "*", < 0.09 ".", > 0.09 "ns") after the model was re-levelled four times to test the difference between accents for each level of Irony and Type. The dotted lines indicate the 95% confidence intervals.

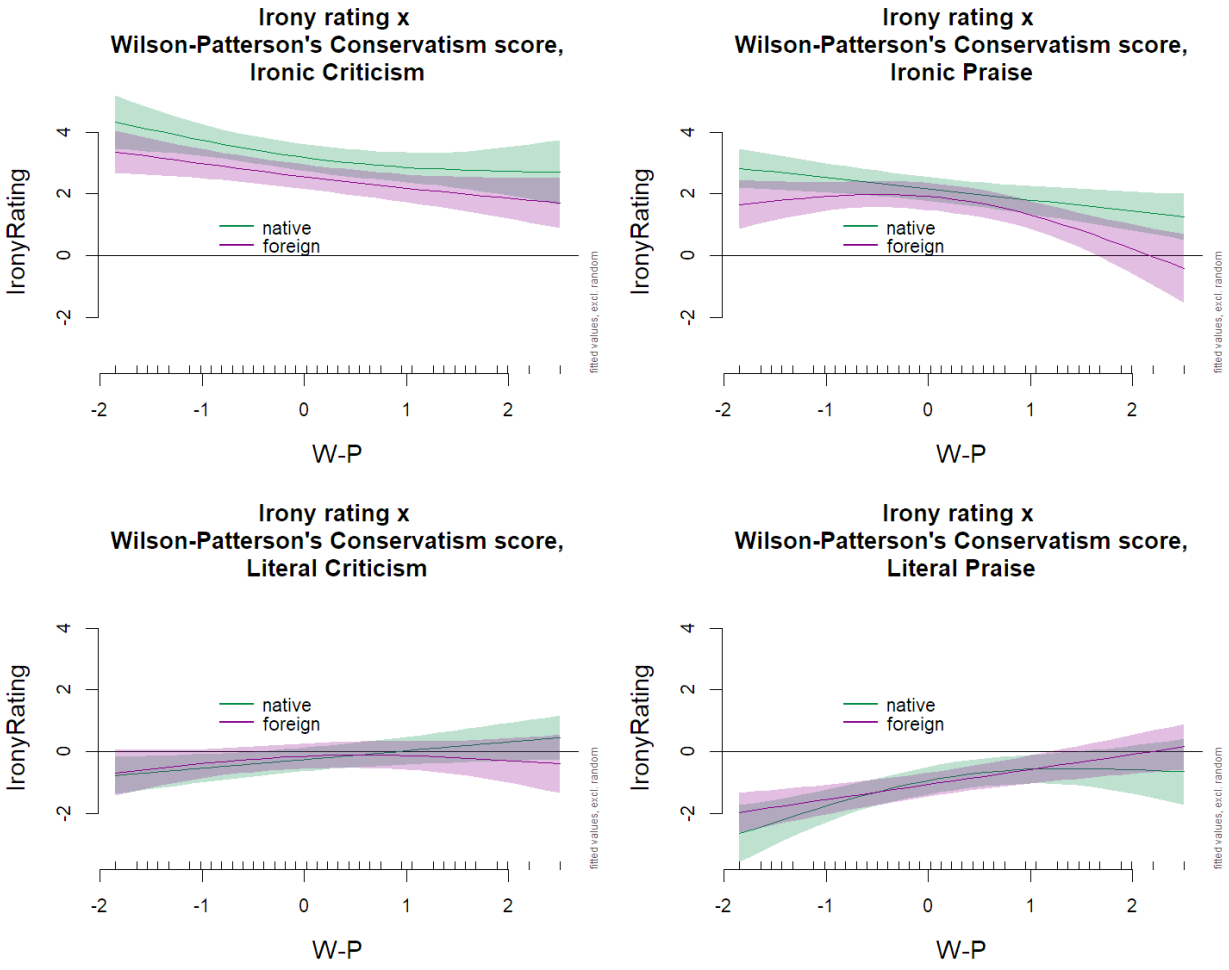

Figure 2.2. Interaction of irony ratings with Wilson-Patterson Conservatism scores. The scores are standardized. Lower scores indicate more liberal orientation.

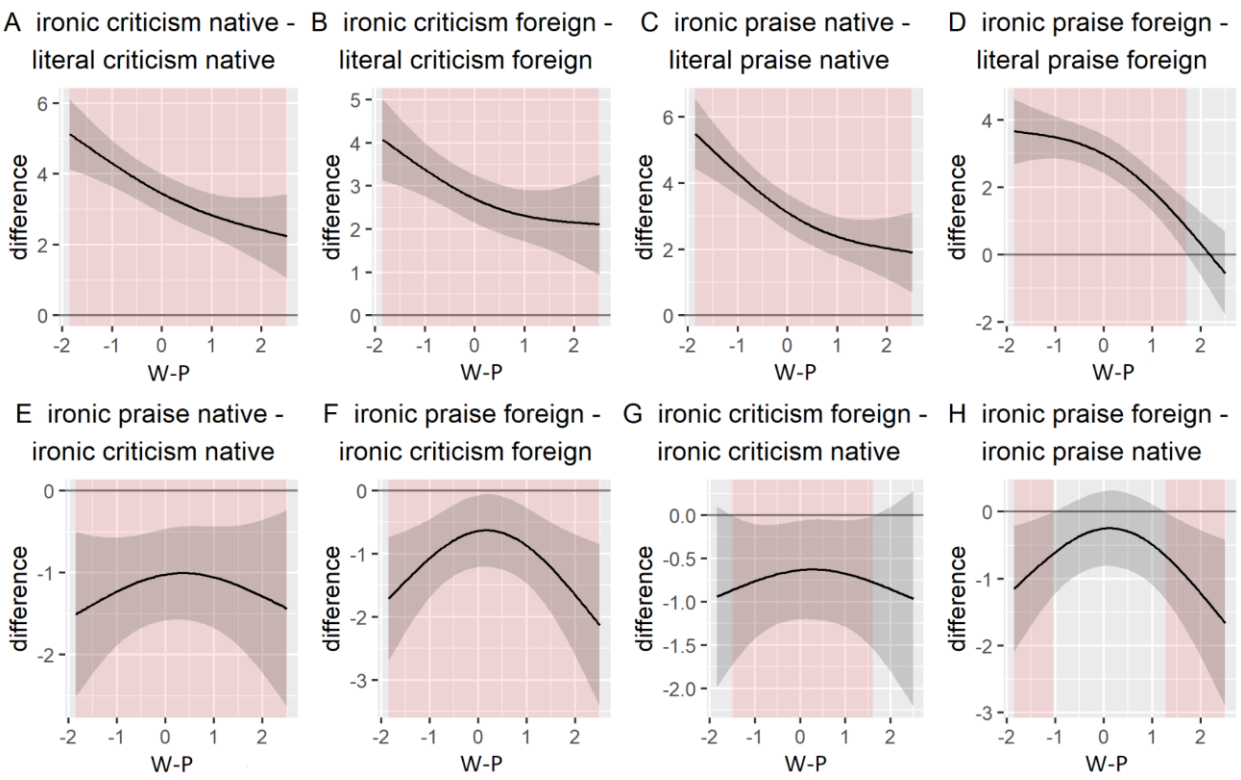

Figure 2.3. Difference plots smoothed by Wilson-Patterson Conservatism scores. Y-axis shows an estimated difference in irony ratings. Shaded areas indicate windows of significant differences.

Certainty rating. Every condition was significantly different from the baseline save literal praise (Table 2.6). This indicates that the difference between the accents was significant for the ironic criticism, with the participants being less certain in their interpretation of foreign-accented utterances. Same as before, we varied the intercept to test for significant differences between the rest of the conditions. No effect of accent was found for either literal criticism or praise (all *p*s > .2). The participants' certainty in the correct interpretation of foreign-accented ironic praise was marginally lower than for native ironic praise (p = .085) (Fig. 2.1B). None of the individual differences improved the model's fit. Once again, we added the intelligibility predictor as a smooth to the final model. It was not significant by itself ($\chi 2$ = 2.23, p = .135) or in interaction with Condition (all *p*s > .6 except for the interaction with literal praise foreign with p = .014). As to the latter model, even though one interaction was significant, the model itself did not have a better fit than the one with Condition only (the difference in REML scores was negligible).

Table 2.6. Summary of the best-fitting GAMM with Certainty Rating as a dependent variable. The model's formula: CertaintyRating ~cond + s(participant, bs = "re") + s(dialog, bs = "re") + s(trial, participant, bs = "fs", m = 1), family = ocat(R = 7). Deviance explained = 22.4%.

Parametric coefficients	**Estimate**	**Std. Error**	**z-value**	**Pr(>| *z* |)**	
(Intercept) (ironic criticism native)	3.98	0.21	18.77	<0.001	***
ironic criticism foreign	-0.56	0.22	-2.61	0.009	**
ironic praise native	-0.60	0.22	-2.74	0.006	**
ironic praise foreign	-0.97	0.22	-4.47	<0.001	***
literal criticism native	-0.44	0.22	-2.02	0.043	*
literal criticism foreign	-0.70	0.22	-3.27	0.001	**
literal praise native	0.11	0.22	0.52	0.606	
literal praise foreign	0.14	0.22	0.65	0.514	
Smooth terms	**edf**	**Ref.df**	**χ2**	**p-value**	
s(participant)	41.05	92.00	69.90	<0.001	***
s(dialog)	82.96	184.00	155.30	<0.001	***
s(trial,participant)	121.44	836.00	2068.60	0.023	**

Appropriateness rating. All praise conditions, literal and ironic, came out significant with respect to native ironic criticism (Table 2.7). Ironic praise was deemed less appropriate than ironic criticism in both accents (Fig. 2.1C). This corroborates the "asymmetry of affect" observed in previous studies (Clark & Gerrig, 1984). Even the fact that our speakers conversed as peers equal in social status apparently did not make ironic praise sound more appropriate. Since we did not specifically test for the effect of social ranks and speakers' relationships, we can only speculate that ironic praise might be judged more appropriate if the speakers are explicitly introduced as close friends. Literal praise, regardless of the accent, was rated more appropriate than the baseline. This suggests that it was successfully conveyed prosodically, without being too similar to sarcastic praise. There was no effect of accent in any condition (all *p*s > .3).

Table 2.7. Summary of the best-fitting GAMM with Appropriateness Rating as a dependent variable. The model's formula: AppropriatenessRating ~cond + s(W-P, by = cond, k = 3, bs = "tp") + s(participant, bs = "re") + s(dialog, bs = "re") + s(trial, participant, bs = "fs", m = 1), family = ocat(R = 7). Deviance explained = 19.3%.

Parametric coefficients	**Estimate**	**Std. Error**	**z-value**	**Pr(>| *z* |)**	
(Intercept) (ironic criticism native)	2.06	0.19	11.00	<0.001	***
ironic criticism foreign	0.03	0.24	0.14	0.893	
ironic praise native	-1.16	0.24	-4.80	<0.001	***
ironic praise foreign	-0.90	0.24	-3.70	<0.001	***
literal criticism native	-0.38	0.24	-1.58	0.114	
literal criticism foreign	-0.12	0.24	-0.50	0.614	
literal praise native	2.01	0.25	8.14	<0.001	***
literal praise foreign	1.73	0.25	7.05	<0.001	***
Smooth terms	**edf**	**Ref.df**	**χ2**	**p-value**	
s(W-P): ironic criticism native	1.58	1.79	2.58	0.346	
s(W-P): ironic criticism foreign	1.00	1.00	5.11	0.023	*
s(W-P): ironic praise native	1.90	1.98	9.10	0.008	**
s(W-P): ironic praise foreign	1.90	1.98	9.24	0.008	**
s(W-P): literal criticism native	1.00	1.00	0.24	0.626	
s(W-P): literal criticism foreign	1.79	1.94	4.66	0.062	.
s(W-P): literal praise native	1.58	1.80	17.85	0.001	**
s(W-P): literal praise foreign	1.73	1.91	19.88	<0.001	***
s(participant)	33.77	91.00	53.95	<0.001	***
s(dialog)	107.25	184.00	262.75	<0.001	***
s(trial,participant)	83.26	835.00	343.89	<0.001	***

Again, only political orientation improved the model's fit. The clearest effect was found for literal praise, which was consistently rated as less appropriate by right-leaning participants compared to left-leaning ones (Fig. 2.4). One explanation could be weaker irony detection aptitude. As discussed above, those who scored higher on the W-P scale were worse at differentiating between ironic and literal comments, which may have led them to believe that literal praise was actually sarcastic. Ratings for ironic praise had the same downward trend in the right end of the scale. The results for criticism were less conclusive. There was a weak trend for left-leaning participants to rate foreign-accented criticism as more appropriate than native, whereas the opposite was true for right-leaning people. This finding requires additional testing. Difference plots produced unclear results (Fig. 2.5). Even though the ironic and literal criticism showed an expected trend, the

window of significant differences was very narrow. The praise conditions showed significant differences across the entire spectrum of W-P scores, but the trend was non-linear (similar to a downward parabola): larger differences at the extreme ends of the W-P spectrum and lower differences in the middle of the scale. Foreign minus native conditions did not show any significant results.

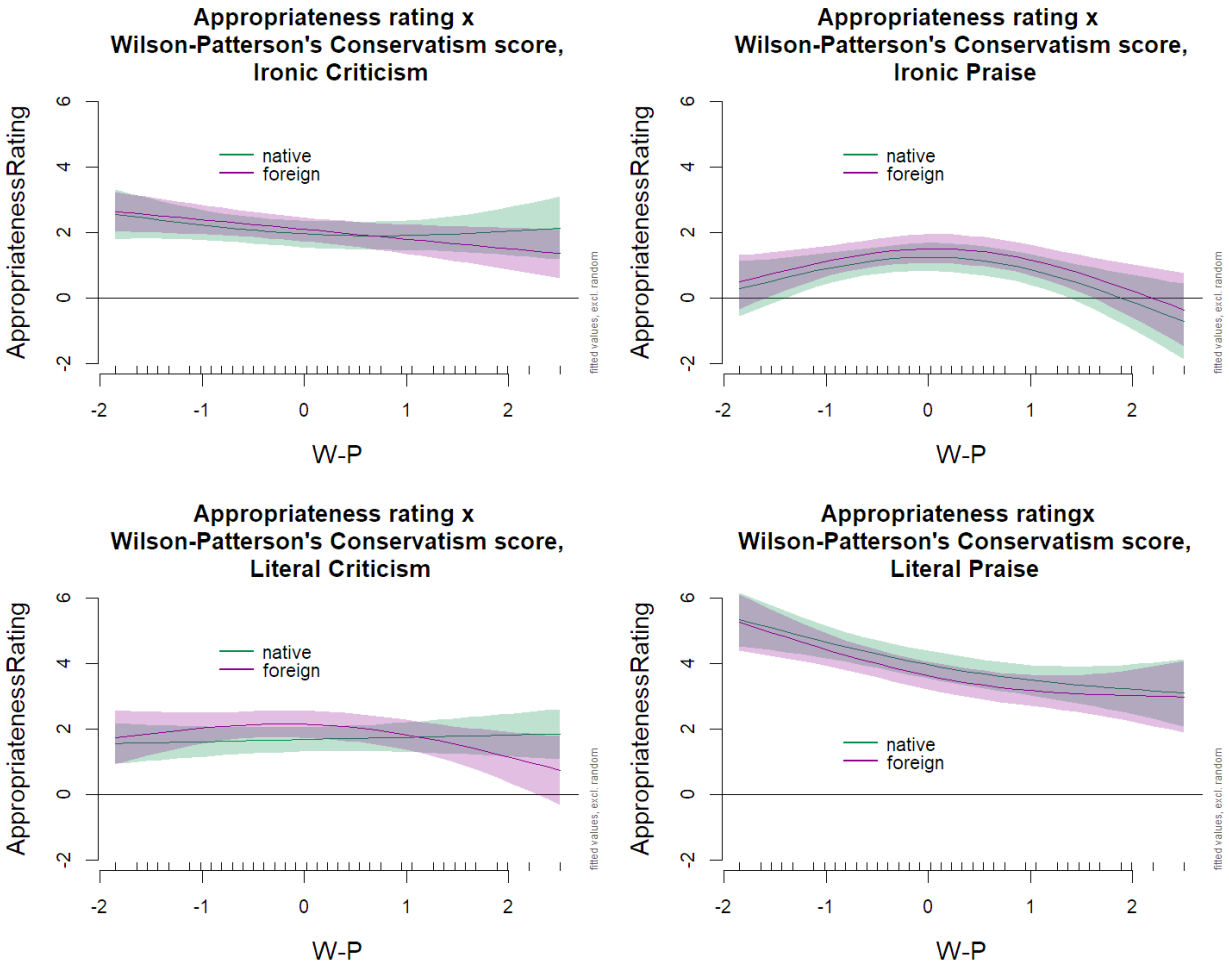

Figure 2.4. Interaction of appropriateness ratings with Wilson-Patterson Conservatism scores. The scores are standardized. Lower scores indicate more liberal orientation.

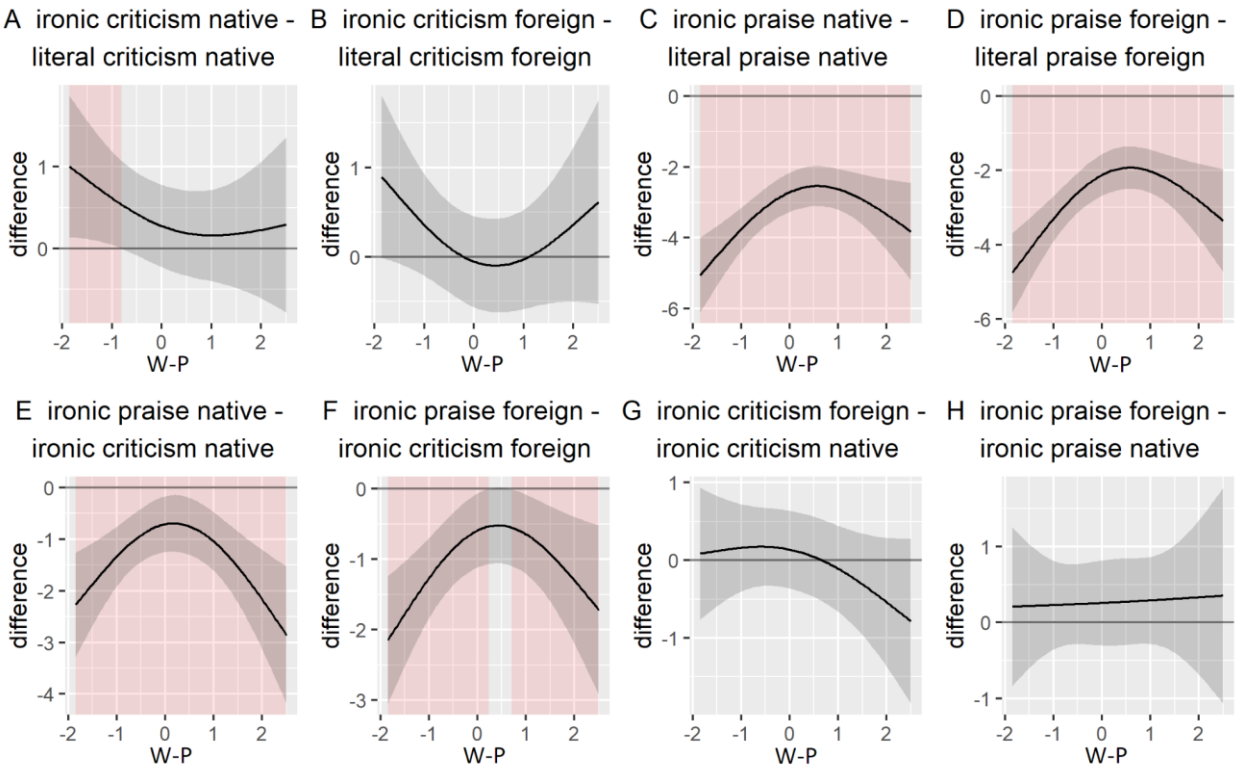

Figure 2.5. Difference plots smoothed by Wilson-Patterson Conservatism scores. Y-axis shows an estimated difference in appropriateness ratings. Shaded areas indicate windows of significant differences.

Offensiveness rating. Both native and foreign-accented literal praise, as well as native literal criticism were significant in the final model (Table 2.8). There was no effect of accent on either ironic criticism or praise (*p*s > .1). Foreign-accented literal criticism was rated less offensive than native criticism ($p = .038$), with no difference for literal praise ($p = .853$). Overall, literal criticism was considered more offensive than ironic criticism in both accents (Fig. 2.1D), which agrees with and extends the findings of Dews, Kaplan, et al. (1995). Finally, native and foreign-accented literal praise was considered less offensive, once again proving that it was successfully delivered prosodically.

Table 2.8. Summary of the best-fitting GAMM with Offensiveness Rating as a dependent variable. OffensivenessRating ~cond + s(participant, bs = "re") + s(dialog, bs = "re") + s(trial, participant, bs = "fs", m = 1), family = ocat(R = 7). Deviance explained = 21.2%.

Parametric coefficients	**Estimate**	**Std. Error**	**z-value**	**Pr(>| *z* |)**	
(Intercept) (ironic criticism native)	1.46	0.19	7.49	<0.001	***
ironic criticism foreign	-0.25	0.25	-1.02	0.308	
ironic praise native	0.06	0.25	0.25	0.804	
ironic praise foreign	-0.34	0.25	-1.34	0.181	
literal criticism native	0.81	0.25	3.25	0.001	**
literal criticism foreign	0.29	0.25	1.18	0.240	
literal praise native	-2.73	0.26	-10.68	<0.001	***
literal praise foreign	-2.69	0.26	-10.50	<0.001	***
Smooth terms	**edf**	**Ref.df**	**χ2**	**p-value**	
s(participant)	36.87	92	61.33	<0.001	***
s(dialog)	112.04	184	298.54	<0.001	***
s(trial,participant)	51.30	836	338.40	0.004	**

2.2.7 Post-tests of the stimuli

To further explore whether some properties of our stimuli could have contributed to the results, we conducted a post-test. We removed the target word (an evaluative adjective) from each dialog frame and asked native speakers of English to fill in the gap with the first word that came to mind. Twenty-three participants completed the task (mean age = 20.2, SD = 4.7) for partial course credit. Since each dialog frame had a positive and a negative version, two lists were created. Additionally, we asked the participants to rate the degree of friendship between the speakers on a 5-point scale (1 = the speakers are total strangers, 5 = the speakers are very close friends).

Since we were not interested in the exact word the participants filled in but rather the valence of the evaluation, no cloze-probability score was computed. Instead, we calculated an "irony score" by dividing the number of ironic completions by all completions. This allowed us to assess whether some contexts were more predictive of ironic continuations than the other. One dialog frame (out of 48) was removed as an outlier because it got 100% of ironic continuations, whereas the next closest score was 36%. We then re-ran the final GAMM model for irony rating with these additional predictors (irony score and friendship score), once again adding them one by one. Both

of them came out significant; however, when added together, only the irony score was significant (p = .025). The model was thus re-fitted with the irony score only and it had a significant improvement over the model without it (REML score -70, p < .001). Fig. 2.6 shows that higher percentage of ironic continuations correlate with higher irony ratings, even though the effect is small. This provides initial motivation for further research to use such a task in a pre-test, since it may potentially explain some variance in the final dataset. Importantly, though, the effects of condition and political ideology remained significant despite the new predictor.

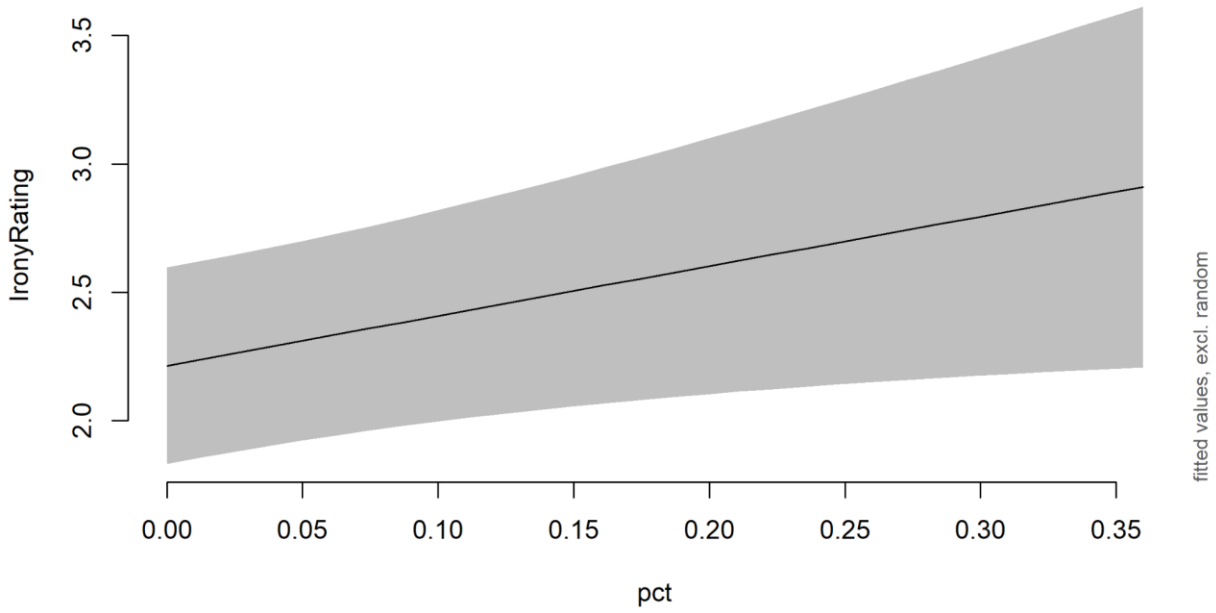

Figure 2.6. A smooth for Irony Score with irony ratings as a dependent variable. X-axis shows the percentage of ironic continuations provided by the participants in the post-test. Y-axis shows an estimated difference in irony ratings.

2.3 Discussion

A tool for social bonding, verbal irony is an essential aspect of social competence. If misunderstood, it may lead to communication breakdowns and injure the relationship. Due to the error-prone nature of foreign-accented speech, listeners might have different expectations when

interacting with non-native speakers and exploit different repair strategies when input violates those expectations. Since irony in its surface form is often contradictory to the preceding context and thus requires complex pragmatic inferencing, the present study tested two hypotheses: 1) foreign-accented irony will be considered less ironic than native irony even when prosodic, contextual, and lexical cues are present, and 2) political ideology, empathy, and the need for cognitive closure will modulate the listeners' perception of foreign-accented irony. To increase the generalizability of our findings, we recorded dialogs with natural prosody and used several rating scales to get a multi-faceted insight into the phenomenon. In brief, we found substantial support for Hypothesis 1 and only partial support for Hypothesis 2. The key findings are as follows:

Both types of irony, criticism and praise, were rated as less ironic when delivered in a foreign compared to a native accent despite multiple cues.

- "The asymmetry of affect" observed in prior research was supported by our data. Overall, the participants considered ironic praise less ironic and less appropriate than criticism.
- Foreign-accented ironic praise predictably evoked most doubts about the speaker's intent (although the effect was marginal) and was judged the least ironic.
- Political ideology significantly affected the participant's irony and appropriateness ratings but not certainty or offensiveness ratings. No effect of either empathy or the need for cognitive closure was found.

Perhaps the most salient finding of this study concerns the different perception of native and foreign-accented irony. These results are not likely to stem from comprehension difficulties for several reasons. First, our stimuli provided rich, supportive context sufficient for disambiguating ironic and literal comments based on semantics alone. Second, the speakers used natural prosody which resulted in longer duration of ironic compared to literal utterances in both accents —the only cross-linguistically consistent cue to irony found to date (González Fuente, 2017) and the one English native speakers use the most (Bryant, 2010). Third, the results of the intelligibility pre-test did not significantly correlate with either irony or certainty ratings. Fourth and most importantly in terms of possible phonetic difficulties, literal comments in both accents were rated equally low for irony, ruling out comprehension issues. As discussed in the Introduction, there may be multiple reasons beyond comprehension for the observed effect—linguistic, social, cultural, or a combination thereof. We will briefly go over each of them below.

From a general linguistic perspective, it has been shown that reanalysis following semantic or grammatical errors appears to be hampered during the processing of non-native speech. Romero-Rivas et al. (2015) found that semantic violations elicited no late positivity when spoken with a foreign accent. There is a sizeable literature attributing this ERP signature to domain-general mechanisms of reanalysis and repair (Kuperberg et al., 2011; Regel et al., 2014; Van Petten & Luka, 2012) or to the integration of multiple sources of information (Friederici, 2017). Importantly, irony has been shown to also systematically elicit late positivity (Regel et al., 2011). Synthesizing all of the above, it appears that the later stage of pragmatic inferencing from ironic comments, be it reintegration of meaning with the information in the long-term memory or repair of disparity between the literal and figurative meanings, proceeds differently for foreign-accented speech compared to native speech. It is not entirely clear whether pragmatic inferencing is just "blocked", reducing cognitive load for the listeners, or whether the attempt at a reanalysis is made but the listener reaches a different conclusion—for instance, that a wrong word was chosen or that a foreign-accented speaker misunderstood the situation. Since we also observed a clear "asymmetry of affect" phenomenon for both accents equally, this pragmatic inferencing appears to be sensitive to the typicality of a trope (ironic criticisms constitute 90% of all ironies according to Dews, Winner, et al., 1995, as cited in Dennis et al., 2001) and its appropriateness (Dews, Kaplan, et al., 1995).

Another potential reason for the results may be less frequent use of irony by non-native speakers, thus producing a surprisal effect and tilting the listeners toward alternative explanations. We are not aware of any research examining the issue of irony use by non-native speakers directly; however, there is converging evidence that non-native speakers are less accurate and slower in irony detection in their non-native language, including Chinese learners of English (Bromberek-Dyzman & Rataj, 2016; Ellis et al., 2021). It is thus possible to hypothesise that, since irony detection seems to be one of the later acquired aspects of L2 pragmatic competence, non-native speakers may use it less often. This needs to be verified by future research.

The perception of social distance and reduced common ground may be another contributing factor. It is possible that the friendly conversational style of our dialogs did not provide enough evidence that the speakers know each other well enough to be ironic. Even though the majority of our items were rated above 3 on a 5-point friendship scale in a post-test, this may not be enough when non-

native speakers are concerned. Less friendly relationship between the speakers has been shown to make irony less prototypical and harder to interpret (Pexman & Zvaigzne, 2004). One may then wonder why this effect was unidirectional. Perhaps foreign-accented speakers are less expected to be comfortable enough to use irony, and this surprisal promotes searching for an alternative explanation.

Cultural stereotypes may also play a role. Regel et al. (2011) showed that extra-linguistic information about the speaker (e.g., the speaker's preferred communicative style, literal or ironic) interacts with pragmatic inferencing in both early and late processing stages. Naturally, it seems plausible that the listeners' stereotypical perception about the frequency and type of irony used by a particular nation can further affect its processing. Most participants in the study by Caffarra et al. (2018) estimated the use of irony as less frequent in Great Britain as compared to Spain. Since we did not ask our participants to provide such an estimation about China and Canada, this explanation remains speculative and needs further validation.

Contrary to what we expected, higher conservatism scores did not correlate with poorer detection of foreign-accented irony—but rather irony as a whole. Right-leaning participants consistently rated ironic comments as less ironic than their left-leaning peers. To our knowledge, no one has previously explored this relationship, so we will offer an explanation building on findings from personality psychology. A meta-analysis by Sibley et al. (2012) suggested that Openness to Experience is the best personality predictor of political ideology, with higher openness predicting higher liberalism. This trait may help left-leaning participants detect playful cues and reject the superficial evaluation uttered by the speaker in favor of a jocular interpretation. Additionally, right-leaning participants rated literal compliments higher for irony and lower for appropriateness than their left-leaning peers. This strongly suggests that they misinterpreted literal praise as sarcastic. Literal praise in our experiment followed explicit sharing of one's achievements (e.g., "*I got a thirty thousand dollar grant for my project!*") which may have been perceived as bragging. Since bragging tends to annoy, it is easy to see how literal praise that followed ("*Oh, your supervisor must be so proud of you right now!*") could be mistaken for sarcasm. Additionally, as shown by Slugoski & Turnbull (1988), literal compliments are likely to be misinterpreted as sarcastic if the speakers dislike each other, whereas literal criticisms are likely to be misinterpreted as ironic compliments if the speakers like each other. Since political ideology had an effect only on literal

compliments and not on literal criticisms, this might suggest that right-leaning participants inferred the "relationship affect" between the speakers incorrectly, erring on the negative side. This idea certainly warrants further investigation.

We did not find any effect of the need for cognitive closure or empathy on any of the ratings. Even though the literature often emphasizes the importance of empathy and mentalizing skills for irony detection, our experiment did not provide any evidence to support this claim. This agrees with the findings of Kieckhäfer et al. (2019) who also failed to find an effect of IRI scores on irony detection beyond personality traits. One possible reason for our findings is the skewness of IRI scores in our sample consisting mostly of young females. This additionally raises the question whether the variance in empathy in healthy individuals is sufficient to lead to noticeable changes in irony detection. Alternatively, the Interpersonal Reactivity Index may not be the right tool for measuring it. It does not seem possible to choose between the above explanations now, and more research is warranted. The need for cognitive closure also did not interact with any of the rating types. Due to the absence of previous irony research using this measure, we can only speculate that irony with sufficient prosody may not be ambiguous enough to trigger affective discomfort, or this affective discomfort may not affect behavioral measures such as ratings.

The results of the current study extend our rather limited knowledge about the processing of foreign-accented speech, specifically as it pertains to pragmatically driven inferences. Our results clearly demonstrate that, when the utterance is counterfactual to the preceding context, the listeners consider the speaker's identity when making an inference about their intention. Further research can explore the timing of these effects using more time-sensitive methods (self-paced listening, EEG). Additionally, future work can try to remove the effects of adverse listening conditions accompanying any foreign-accented speech from the picture and use written stimuli cuing the speaker's identity.

Naturally, our study had limitations. First, the convenience sample resulted in somewhat skewed distributions of individual difference scores. Second, due to a prohibition on in-person testing as a result of the COVID-19 pandemic, we had to transfer our experiment online. Even though we emphasized the importance of completing it in one sitting and without external distractions, we cannot be certain that our recommendations were respected. Third, even though accent did not interact with any condition when acoustic features such as pitch, duration, and speech rate were

analyzed, it remains possible that this interaction was present on some other level of phonological analysis (for instance, intonation). And finally, as discussed in the Data analysis section, opting for generalized additive mixed modelling meant that we had to represent our three predictors with two levels each as one grouping predictor with eight levels. Even though we reran the model for irony ratings without individual differences as a 2 × 2 × 2 and the results converged, this should still be noted as a potential limitation.

In conclusion, we showed that even an intelligible foreign accent affects irony comprehension, and that irony detection skills are in turn affected by political ideology. Our results are based on a novel statistical method for analyzing ordinal data (generalized additive mixed modelling) which does not require treating ordinal data as simple integer-valued. Most importantly, we used dialogs that contained multiple cues to irony (supportive context, prosody, and intensifying lexical markers biasing an ironic interpretation), which makes the data ecologically valid and more readily generalizable to language processing in the real world. The results of this study demonstrate the importance of taking interpersonal differences in language processing into account rather than averaging over them, in particular when it comes to irony detection. This study also adds to the growing body of evidence that nonnative speakers face numerous challenges in day-to-day communication, which may eventually translate into negative consequences for many aspects of life. Even though a few studies showed unexpected social advantages for being a non-native speaker (Fairchild et al., 2020; M. Ip & Papafragou, 2021), most work still converges on negative social attitudes toward non-native speech (Bresnahan et al., 2002; Fraser & Kelly, 2012; Fuertes et al., 2012; Gluszek & Dovidio, 2010; Lev-Ari, 2015; Lev-Ari & Keysar, 2010). Last but not least, the results of this study extend the previous findings obtained in a more controlled lab setting to an online format, suggesting that in-lab and online data collection can deliver converging results.

2.4 Ethics

All participants gave consent to participating in the experiment after they were provided with a full description of the study and were at liberty to withdraw from the experiment at any point for any reason with no consequences. The plan for this study was reviewed for its adherence to ethical guidelines by a Research Ethics Board at the University of Alberta (reference number Pro00102750).

2.5 Acknowledgments

We thank Hesheng Xing, Scott Perry, Tyler Schnoor, and Wenfu Bao for their kind help with the stimuli. We are also grateful to Anja Arnhold for providing PRAAT scripts for the acoustic analysis of the stimuli. We sincerely thank the editor and three reviewers for their insightful comments and suggestions which helped us to improve this manuscript. This research was supported by a Social Sciences and Humanities Research Council of Canada Partnership Grant (Words in the World, 895-20161008).

Chapter 3

Acoustic correlates of irony in native and Mandarin-accented English

Abstract

Whether irony in English is associated with any distinct prosodic correlates apart from longer duration has been a subject of intensive research, yet little consistency has been found to date. Virtually all existing studies focus on prosodic parameters calculated over entire utterances, with little attention to levels such as vowels or individual words, and use ironic stimuli of very different length and syntactic structure. Importantly, irony in English is more and more commonly expressed by not only native but also non-native speakers. To date, no data exists about prosodic contrasts in foreign-accented irony. To fill in these gaps, we did a fine-grained acoustic analysis of native English and Mandarin-accented ironic and literal speech used as stimuli in our previous study (Puhacheuskaya, V., & Järvikivi, J. (2022). *I was being sarcastic!*: The effect of foreign accent and political ideology on irony (mis)understanding. *Acta Psychologica*, 222, 103479. https://doi.org/10.1016/j.actpsy.2021.103479). Two types of irony were examined, ironic praise and ironic criticism, and compared to their literal counterparts. We hypothesized that foreign-accented irony would be more prosodically marked to minimize any risk of misunderstanding. We also predicted that ironic praise would be more marked than criticism because the risks of listeners taking it at face value should outweigh any benefits offered by being deliberately ambiguous. One hundred and ninety-two utterances were analyzed at different levels of granularity—vowels, orthographic words, multi-word intervals, and whole utterances. In agreement with previous research, we found that irony was indeed associated with several prosodic correlates (e.g., longer duration, higher intensity, higher f0 variability, and a larger f0 range), although not all of them were significant at every level and in every interval. Most importantly, accent interacted with irony and type (criticism or praise) at multiple levels of analysis. Consistent with our prediction, foreign-accented irony differed from literal speech significantly more than native irony did. In addition,

we also found that foreign-accented praise was marked significantly more than criticism in at least one measure: mean vowel duration. Possible explanations for the observed effects and directions for future research are discussed.

Keywords: irony, ironic praise, ironic criticism, foreign accent, acoustic analysis, prosody, acoustic cues

3.1 Introduction

Whether there is an "ironic tone of voice" has been a matter of considerable debate, and the evidence is far from conclusive (Anolli et al., 2002; Attardo, 2000, 2001; Attardo et al., 2003; Bryant, 2010, 2012; Bryant & Fox Tree, 2005; Cheang & Pell, 2008; González Fuente, 2017; Rockwell, 2000, 2007). On the one hand, having an infallible auditory or written marker of irony would deprive it of two important social functions, retractability and building in-group solidarity. As pointed out in Attardo (2000) and Brown & Levinson (1987), ironic ambiguity allows one to withdraw their statement should the speaker take extreme offense and avoid sanctions that would otherwise follow. In addition, shared irony creates a feeling of belonging to an in-group (R. T. Lakoff, 1990). As Gretchen McCulloch puts it in her book *Because Internet*, "Irony is a linguistic trust fall. When I write or speak with a double meaning, I fall backwards, hoping that you'll be there to catch me. The risks are high: misaimed irony can gravely injure the conversation. But the rewards are high, too: the sublime joy of feeling purely understood, the comfort of knowing someone's on your side" (McCulloch, 2019, p. 161). In other words, irony serves as a means to emotionally connect with people who decoded your indirect meaning, thus strengthening an in-group feeling. On the other hand, not marking irony in any way bears its own risks, even more so when it comes to ironic compliments.[6] In contrast to ironic insults that allow the speaker to avoid being impolite and save face ("*What a brilliant idea!*" said of an idea that is far from brilliance), ironic compliments do the exact opposite. Being mock negative evaluations said in the face of positive circumstances ("*What a dumb idea!*" said of a great suggestion), ironic compliments are usually intended to tease and be playful. However, since they violate conversational etiquette and politeness norms, they evoke a more negative affect compared to ironic insults (Clark & Gerrig,

[6] We will use "compliments"/"praise" and "insults"/"criticism" interchangeably throughout the article.

1984) and have a larger social cost. In terms of prosody, it would thus make more sense to mark them overtly, or at least more strongly than ironic insults. To our knowledge, there is only one study that examined prosodic correlates of ironic insults and ironic compliments in English (Mauchand et al., 2018). Surprisingly, though, none of the acoustic measures significantly correlated with ironic compliments (teasing), whereas ironic insults could be distinguished based on slower duration and reduced pitch variability. It should be noted that the authors only looked at global measures across entire utterances (duration, mean f0, f0 SD, intensity variation, etc.) and report multiple discrepancies across their stimuli, which merits further research.

What is more, there is substantial cross-linguistic variation in acoustic correlates of irony, with slower duration being the only consistent cue across different languages identified so far (see González-Fuente, 2017 for an overview). Bryant (2010) suggested that this feature serves a cognitive function, giving the listeners extra time needed to reject the literal meaning and arrive at an ironic interpretation.[7] More importantly for our purpose, this cross-linguistic variation in ironic prosody presents a challenge for non-native speakers. About seventy percent of the 1.4 billion speakers of English in 2021 speak it non-natively according to Ethnologue (Eberhard et al., 2021), which makes foreign-accented irony a phenomenon of immediate interest. Only a few studies so far have examined the perception of foreign-accented irony (Caffarra et al., 2018, 2019; Puhacheuskaya & Järvikivi, 2022), and none of them did a fine-grained acoustic comparison of native and non-native irony markers. Moreover, Caffarra et al. (2018, 2019) explicitly instructed their native and non-native speakers to maintain flat intonation and slow speech rate for all ironic and non-ironic utterances, making such an analysis meaningless. Our previous study (Puhacheuskaya & Järvikivi, 2022) addressed these limitations by using natural prosody but still only reported a few global measures (duration, pitch, and speech rate) calculated over entire utterances. While no interactions between irony and accent were found, it remains possible that they were present at some other level or in some other measures. Notably, since foreign-accented irony was judged significantly less ironic compared to native one in that experiment, it becomes

[7] However, see Gibbs Jr. (2002) for an account that does not require access to literal meaning for irony interpretation and Kowatch et al. (2013) for empirical findings that undermine the idea of longer processing times for ironic utterances.

particularly important to know whether some of these effects stem from different prosodic contrasts in non-native ironic speech.

Since non-native speakers consistently face negative perception in most areas of life (Foucart et al., 2019; Fraser & Kelly, 2012; Fuertes et al., 2012; Lev-Ari & Keysar, 2010), and since they, in fact, know of and *expect* stigmatization from others on the grounds of their accent (Derwing, 2003; Gluszek & Dovidio, 2010), it seems likely that they would prosodically mark ironic utterances to a larger extent than native speakers to avoid misunderstanding. This should be particularly true for ironic compliments, since non-native speakers may try to avoid language that would be perceived as rude and exacerbate this anticipated negative perception of themselves. Previous studies called this asymmetry between the perception of ironic criticism and compliments "the asymmetry of affect" (Clark & Gerrig, 1984), and it largely stems from the fact that being impolite, even out of mere play, violates implicit conversational norms. Extensive prosodic marking may thus increase the chances that the comment will be interpreted as friendly teasing and help avoid this negative affect.

Since our dataset included native and Mandarin-accented English, we will review major studies that investigated prosodic correlates of irony in English and Mandarin Chinese before we proceed to describing the goal and research questions of this paper in more detail.

3.1.1 Ironic prosody in English

Several perception experiments showed that native speakers of English use prosody to disambiguate between literal and ironic meanings, and that they can do so reliably even in utterances taken out of context (Mauchand et al., 2019; Rockwell, 2000). Similar findings were observed for Spanish: In Padilla (2011), as quoted in González-Fuente (2017, p. 44), 49 percent of listeners considered the tone of voice more important than context in making a decision, 50 percent considered them equally important, and only 2 percent put context above prosody. This testifies to the importance of prosody in conveying the ironic intent.

Despite that, production experiments found a lack of systematicity in how English speakers prosodically contrast ironic utterances to the preceding context (Anolli et al., 2002; Rockwell, 2000). Specifically, Anolli et al. (2002) showed that prosodic marking of irony depended on the type of context in which it was embedded. For instance, cooperation contexts evoked irony of high

intensity with either a high and variable pitch or a low and monotonous pitch. Conflict contexts were associated with slow speech rate with either a high and variable pitch plus high intensity or a low and not variable pitch plus low intensity. Importantly, though, studies on ironic prosody are hard to compare, because they have been done on scripted and spontaneous[8] verbal irony, on stimuli recorded by professional actors and naïve subjects, using acoustic measurements calculated with technological devices or perceptual coding by human coders, with highly variable stimuli in terms of syntactic structure, on different subtypes of irony (kind, sarcastic, etc.) and even with different control conditions (e.g., neutral or ambiguous prosody). Naturally, with so much variation, it is hard to draw strong conclusions about the "ironic tone of voice" in English, the very notion of which is likely oversimplified (Bryant, 2012; Bryant & Fox Tree, 2005). Overall, irony is most often associated with some sort of an "emphatic stress of suprasegmental features" (pitch, intensity, and rate of articulation, among others) which makes it sound like it "plays with the voice" (Anolli et al., 2000, p. 300). We will now review the main acoustic markers that have been found in previous research to be more or less consistently associated with irony.

Most of the studies have found differences in pitch between ironic and literal utterances, with the former being produced with either a higher (Anolli et al., 2000; Rockwell, 2000) or a lower pitch (Cheang & Pell, 2008); more (Bryant, 2010) or less (Cheang & Pell, 2008) f0 variability; greater absolute pitch changes (Bryant, 2010); and a reduced/flattened (Milosky & Wrobleski, 1994); as quoted in (Rockwell, 2007) or increased (Haiman, 1998) pitch range. In addition, Haiman (1998, p. 31) noted an "inverse pitch obtrusion", where a stressed syllable in ironic utterances is uttered with a lower pitch than the surrounding syllables. Attardo et al. (2003) also note a strong within-statement pitch contrast: "usually a large pitch range in the first part of the utterance, followed by a phrase with a highly compressed pitch range (though in several examples this pattern is reversed)" (p. 247). Some authors also discuss a flat intonation contour associated with irony (neither rising, nor falling) (Milosky & Wrobleski, 1994; as quoted in Attardo et al., 2003), question intonation (Attardo, 2001), or "exaggerated intonation patterns" (e.g., a singsong melody or falsetto) (Attardo, 2001).

[8] So far studies designed to elicit sarcastic responses naturally had very limited success (see Rockwell & Theriot, 2001).

Intensity/amplitude/energy has also sometimes been found as a prosodic correlate of irony. Ironic utterances are usually associated with higher intensity (Anolli et al., 2000; Rockwell, 2000) and lower amplitude variability (Bryant, 2010; Bryant & Fox Tree, 2005).

As discussed above, the most consistent cue to irony has been found to be duration. Ironic utterances are usually associated with a larger mean duration (Anolli et al., 2000; Bryant, 2010; Rockwell, 2000), slower rate of articulation (Anolli et al., 2000; Cutler, 1974), and increased length of vowels (Anolli et al., 2000; Haiman, 1998). The data on the duration of pauses between words is inconsistent: Anolli et al. (2000) found shorter pauses in ironic utterances whereas Haiman (1998) noted extra long, "heavy" pauses (p. 39).

Other miscellaneous cues to irony identified in prior research include nasalization (Cutler, 1974; Haiman, 1998), lower harmonics-to-noise (HTN) ratio (Cheang & Pell, 2008), and pronounced pitch accents, often on "multiple syllables of the same word" (p. 250) (see also "exaggerated stress" in Haiman 1998, p. 39). Among more obvious cues to irony are also laughter syllables across the target utterance or even before or after it (Haiman, 1998).

Summarizing, a wide variety of experimental designs and techniques used to analyze the data are likely responsible for a lack of consistent patterns in ironic markers and sometimes directly contradictory results (such as lower or higher pitch found for ironic utterances). One could hypothesise that the precise nature of marking does not matter as much as the fact that the ironic utterances are simply *contrasted* with the preceding context. However, even that is not consistently found, and some studies failed to find any other markers apart from a slower duration (see Bryant, 2010). Why duration seems to be so important that it is found across many different languages is not clear yet and may, as suggested by Bryant, have a cognitive function, such as giving the listener some extra processing time. Apart from duration, it is highly likely that different types of irony or different contexts in which it is embedded are associated with their own prosodic markers. In addition, verbal irony in everyday communication is most often paired with a variety of facial features or even gestures, which may express whatever is missing in the auditory stream.

3.1.2 Ironic prosody in Mandarin Chinese

Research on prosodic correlates of irony in Chinese is very scarce, even more so when it comes to Mandarin. We were able to find only two studies on irony in Mandarin, one of which examined

the prosody of ironic insults (S. Li et al., 2020) and the other the prosody of ironic compliments (S. Li & Gu, 2021), a subtype that has not been sufficiently researched even in English. The authors report that ironic comments constitute about 13.5 percent of all comments in daily interactions in Chinese, which makes irony more frequent in Chinese compared to the 8 percent reported for English (Gibbs Jr., 2000). The first study (S. Li et al., 2020) found that ironic insults in Mandarin were associated with a particular voice quality: creakier voice. This voice quality was characterized by several prosodic measures, such as lower f0, a higher contact quotient, and a higher HNR. The second study (S. Li & Gu, 2021) compared ironic compliments to literal insults, without looking at ironic insults or literal compliments. An acoustic analysis found significant gender differences in prosodic cues to ironic compliments. For instance, while female speakers used a lower mean f0 with a smaller SD to mark ironic compliments (for both target and non-target words), male speakers had a higher mean f0 for both targets and non-targets and a larger f0 SD in targets only. This testifies to an important role of gender differences in ironic prosody. Slow duration was also a significant predictor for ironic compliments, similar to what has been found for ironic insults in English and other languages (Anolli et al., 2000; Bryant, 2010; González Fuente, 2017; Rockwell, 2000). Additionally, the authors reported a lower mean intensity with a smaller SD for ironic compliments. Interestingly, they also found that ironic compliments were associated with a breathier voice (as derived from a higher H1-H2, a higher HNR and a lower contact quotient), which contrasts them with a creaky voice the authors found for ironic insults (S. Li et al., 2020). A higher HNR and a lower contact quotient were significant for ironic compliments for both genders, but the effect size was larger for men.

All in all, research done so far has identified two distinct voice qualities associated with ironic insults and ironic compliments in Mandarin. In addition, several interactions of prosody with gender have been found, with women using a lower mean f0 and men a higher mean f0 to mark ironic compliments. Longer duration of ironic utterances found in Mandarin Chinese agrees with the data reported for English in the previous section. As far as mean f0 and f0 SD go, these results are compatible with some studies on English and disagree with others. Whether the same prosodic patterns would be observed for Mandarin-accented irony in English remains to be seen and was thus one of the research questions of this paper.

3.1.3 The present study

The purpose of this paper was to identify acoustic markers of ironic insults (sarcasm) and ironic compliments (teasing) as compared to their literal counterparts in native and Mandarin-accented English. To the best of our knowledge, this is the first study to do a fine-grained acoustic analysis of foreign-accented irony. Specifically, we wanted to look at different levels of granularity (vowels, syllables, words, multi-word intervals, and whole utterances) to check whether foreign accents interacted with irony in a more complicated way than the results of our previous coarse-grained analysis suggested (Puhacheuskaya & Järvikivi, 2022). In addition, we wanted to investigate whether ironic praise and ironic insults are associated with distinct prosodic patterns. We hypothesized that ironic praise would be more prosodically marked compared to ironic insults because the possibility that the speaker will take it at face value has a larger social cost and offers fewer benefits. We also expected foreign-accented irony to be more prosodically contrasted with literal speech than native irony. Finally, due to the reasons described in the Introduction, we expected foreign-accented ironic praise to be the most prosodically marked among all conditions.

3.2 Materials and methods

3.2.1 Materials

The data analysed in this study was originally recorded for Puhacheuskaya & Järvikivi (2022). The speakers were seated in two sound-treated WhisperRoom booths and used Countryman Earset microphones. The audio was saved in the WAV format with a 44.1 kHz sampling rate and a 16-bit precision. There were four speakers in total: two native speakers of Canadian English (from Ontario and Alberta) and two native speakers of Mandarin Chinese (from Mainland China). All the speakers were male and in their mid-twenties. The stimuli were recorded as dialogs, and the order in which the speakers started reading the lines was counterbalanced to reduce mimicking behaviours. The crucial characteristic of the stimuli was their natural prosody. The speakers were instructed to read ironic utterances with the intonation they would normally use to convey irony in day-to-day communication. In addition, all target utterances were embedded in rich contexts that should have facilitated natural prosody. While prosody may be a less reliable cue when coming from non-native speakers, the main idea was to test the generalizability of previous studies which

used an artificially flat intonation and a slow speech rate (Caffarra et al., 2018, 2019) to a more ecologically-valid setting. A total of one hundred and ninety-two dialogues were used in the main experiment. Each native speaker was paired with both foreign-accented speakers so that each pair of speakers read a quarter of the stimuli (forty-eight dialogues). Two within-speaker factors were Irony (ironic, literal) and Type (praise, criticism). Accent (foreign, native) was a between-speaker factor. The native and non-native speakers in each pair alternated between dialogues with respect to speaking the target and non-target utterances. Thus, half of all the targets were in native accents (ninety-six) and the other half in the foreign accent. Full description of the stimuli and the recording procedure are provided in Puhacheuskaya & Järvikivi (2022) in the Materials section.

A Praat script used for segmentation and annotation together with the segmented audio is available on OSF. For acoustic measurements (mean duration, mean intensity, min and max f0 as well as f0 at 10 equidistant points per interval, etc.) we used Anja Arnhold's script available at https://sites.ualberta.ca/~arnhold/praatScripts.html. The intensity of the stimuli was uniformly scaled to 70 dB. Only the target utterances were analysed, which were always the last utterances of the dialogues. An example dialogue in all four (within-speaker) conditions is provided in Table 3.1.

Since the stimuli were not specifically designed for a detailed acoustic analysis, some methodological issues were inevitable, for instance the occasional presence of phones that cannot be reliably segmented (central and lateral approximants, weak voiced fricatives, etc.) in target words (Turk et al., 2006). Additionally, all target utterances varied in length and syntactic complexity. We segmented and annotated the utterances at four granularity levels: vowels, syllables, orthographic words, and EEG intervals. An EEG level was added because the stimuli were initially designed for an EEG experiment, which eventually had to be turned into a rating study due to a COVID-related prohibition on in-person testing. All EEG tiers consisted of five intervals: #1 (beginning), #2 (pre-target), #3 (target), #4 (spillover), and #5 (wrap-up). Since each utterance had the same EEG structure, with the third interval always being the target word that disambiguated between an ironic and literal interpretation, this allowed for a more direct comparison of utterances. The boundaries for all intervals in all tiers were identified by both visual inspection of the spectrogram and listening to the audio. Several criteria were used to help with visual identification, such as the onset and offset of frication noise for fricatives, a release burst

followed by a cessation of voicing for plosives (if two plosives occurred next to each other, the silence between them was split evenly), and formant movements for approximants and vowels (Turk et al., 2006). If present, aspiration noise was separated from the onset of vowels for the vowel tier. For the vowel tier, /ɹ/ in coda position was included in the preceding vowel. In addition to segmentation, pitch objects were automatically generated and then manually corrected by both visual inspection of the spectrogram and listening to the audio. When words were realized in creaky voice, which was especially common in the utterance-final position, the pitch could not be reliably computed and thus the corresponding part of the pitch object was manually silenced.

Table 3.1. The list of experimental conditions with examples. Every dialog in this table was recorded twice, with the native and foreign-accented speakers swapping roles. The order in which the speakers started reading the dialogs was counterbalanced. Target words are in bold.

Irony Type	Ironic	Literal
Praise	—How did the dinner with your wife's parents go? —I spent the whole day cooking salmon and making drinks. They finished everything, down to the last crumb, and asked for seconds. —You surely are the most **terrible** cook in this part of town	—How did the dinner with your wife's parents go? —I spent the whole day cooking salmon and making drinks. They finished everything, down to the last crumb, and asked for seconds. —You surely are the most **talented** cook in this part of town
Criticism	—How was your Saturday dinner with your wife? —I decided to try a new fish recipe. I probably messed up because she couldn't finish it and the next day got really bad food poisoning. —You surely are the most **talented** cook in this part of town	—How was your Saturday dinner with your wife? —I decided to try a new fish recipe. I probably messed up because she couldn't finish it and the next day got really bad food poisoning. —You surely are the most **terrible** cook in this part of town

3.2.2 Data analysis

The analyses were done in R (4.1.1) (R Core Team, 2020) and are also available on OSF. Linear-mixed models for each dependent variable were run using the *nlme* (3.1-152) (Pinheiro et al., 2021) and *buildmer* (2.2) (Voeten, 2021) packages. *Buildmer* performs stepwise model elimination based on the chosen criterion, such as the change in log-likelihood or its significance, Akaike's

Information Criterion, Bayesian Information Criterion, and other. We did backward model selection based on LRT (the likelihood ratio test), starting with a fully specified model (fixed effects of Irony, Type, Accent and all interactions between them as well as two random intercepts of Dialog and Speaker). Random slopes were also tested but the models including them did not converge. Reference levels were taken alphabetically: *ironic*, *criticism*, and *foreign*, respectively. Sometimes the model suggested as the best by *Buildmer* turned out to have insignificant predictors when fitted afterwards using the *lme*() function. In those cases, we took a conservative approach (due to a small dataset) and removed predictors until all that remained in the model were either significant or marginally significant. For post hoc comparisons we used the *emmean* package (1.7.1-1) (Lenth et al., 2023) with a Bonferroni adjustment for multiple comparisons.

We analysed eight prosodic parameters specified below:

- Mean vowel duration: for utterances
- Mean f0: for utterances
- F0 SD: for utterances
- F0 range: for utterances
- Intensity SD: for utterances
- Intensity range: for utterances
- Mean intensity: for utterances, EEG intervals, and stressed vowels in the target interval
- Mean duration: for EEG intervals and stressed vowels in the target interval
- Max f0: for EEG intervals and stressed vowels in the target interval

There were 192 data points for utterance-level statistics, 192±8[9] data points for EEG interval-level statistics (see section 3.2.1 for an explanation) (per interval; 960±10 in total), and 192 data points for stressed vowels in the target interval. The procedure for calculating acoustic measures over entire utterances was as follows:

Mean f0 was calculated by first calculating mean f0 over every vowel in the utterance (taking f0 at 10 equidistant points per vowel as the basis) and then calculating the mean of all vowel means. F0 SD was calculated in the same way except for calculating SD instead of the mean over all vowel means. For the f0 range, we subtracted min f0 from max f0 (measured by the Praat script) per each

[9] This variance was due to missing f0 measures for some of the intervals.

utterance. Mean intensity was calculated for each vowel automatically by the Praat script, after which we calculated the mean of all vowel means. The procedure for calculating intensity SD and range was the same as for f0. Mean vowel duration in an utterance was calculated as a mean over all vowel means.

3.2.3 Results

Utterance-level statistics. Since our previous analysis (Puhacheuskaya & Järvikivi, 2022) has already found ironic utterances to have longer mean duration, we did not calculate the same parameter again here. Instead, however, as previous research reported an increased length of vowels in ironic utterances (Anolli et al., 2000; Haiman, 1998), we tested whether mean vowel duration differed between ironic and literal conditions. Linear-mixed modelling showed a main effect of Type ($t = 4.09$, $p < 0.001$), three significant two-way interactions Irony × Type ($t = -3.52$, $p < 0.001$), Irony × Accent ($t = -2.84$, $p = 0.005$), Type × Accent ($t = -2.68$, $p = 0.008$), and a significant three-way interaction Irony × Accent × Type ($t = 2.31$, $p = 0.022$). The duration of vowels was consistently larger in native ironic compared to literal conditions (ironic criticism vs. literal criticism: t.ratio = -3.63 , $p = 0.004$; ironic praise vs. literal praise: t.ratio = -3.99, $p = 0.001$) (Fig. 3.1, right panel), but the difference between criticism and praise themselves was not significant in either the ironic or the literal conditions (all ps > 1). In contrast, foreign-accented ironic praise had a much larger mean vowel duration than ironic criticism (t.ratio = 4.10, $p < 0.001$) (Fig. 3.1, left panel). This corroborates our prediction that foreign-accented ironic compliments would be the most marked among all conditions.

For f0 mean, there was a significant main effect of Irony ($t = -3.53$, $p < 0.001$) and a significant two-way interaction between Irony and Accent ($t = 3.09$, $p = 0.002$). As shown in Fig. 3.2, the effect of Irony was meaningless due to the presence of a higher-level interaction. Foreign-accented utterances had a higher mean f0 in the ironic compared to literal condition (t.ratio = -3.53, $p = 0.002$), while there was virtually no difference between the two for native speech ($p > 0.8$).

For f0 SD, Irony was the only significant predictor ($t = -4.36$, $p < 0.001$), with ironic utterances having higher f0 variability than literal utterances. No effect of Accent or Type were found.

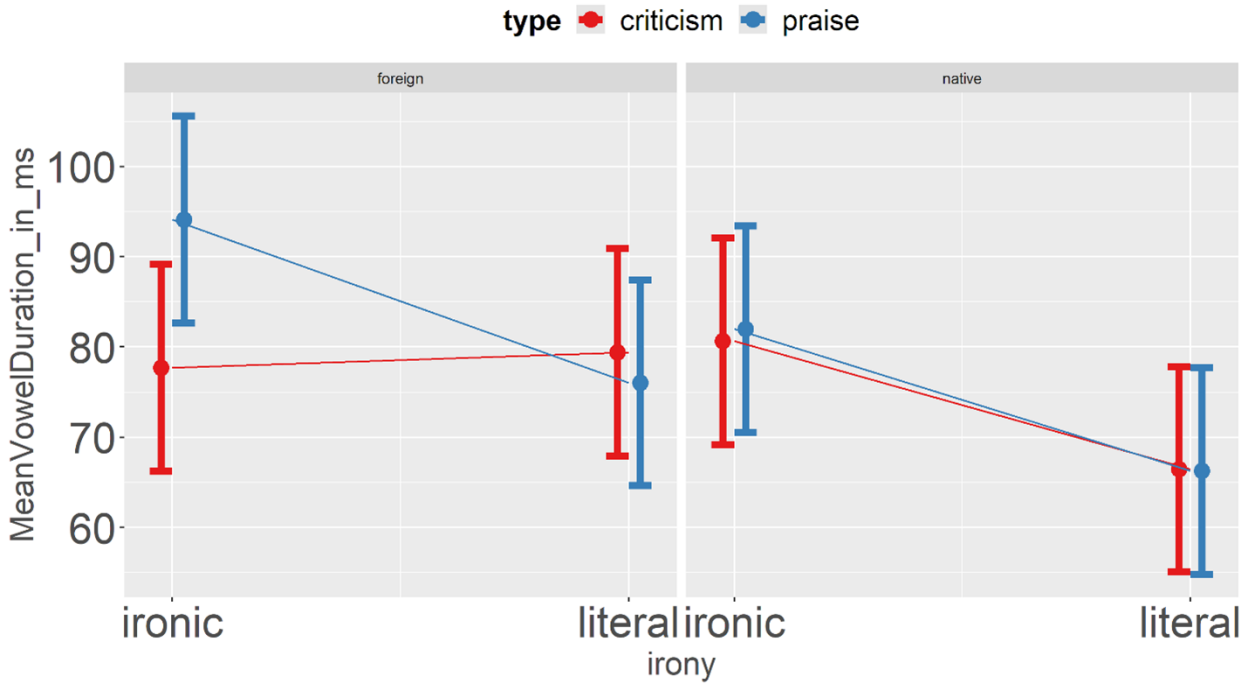

Figure 3.1. Predicted values (marginal effects) of the interaction between Irony, Type, and Accent with mean vowel duration as a dependent variable.

For f0 range, there were main effects of Irony (t = -4.80, p < 0.001) and Type (t = 3.44, p < 0.001) and two significant two-way interactions, Type × Accent (t = -3.51, p < 0.001) and Irony × Accent (t = 2.72, p = 0.007). The fixed effect of Irony was meaningful in the presence of a higher-level interaction, with ironic utterances having a larger f0 range than literal ones (see Fig. 3.3A). Post-hoc comparisons showed that only the difference between foreign-accented ironic and literal utterances was significant (t.ratio = -4.79, p < 0.001), whereas the difference between native ironic and literal utterances was not. In addition, the difference between praise and criticism was only significant for the foreign accent (t.ratio = 3.44, p = 0.003) (Fig. 3.3B).

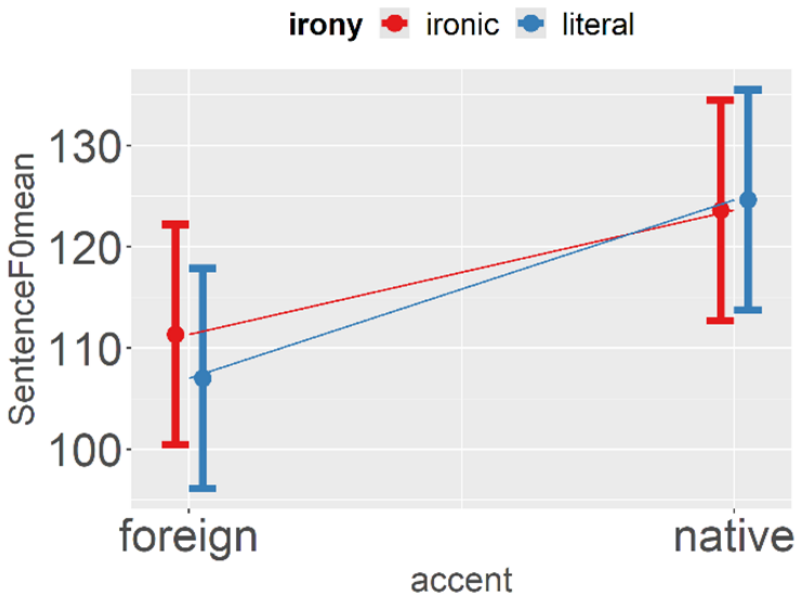

Figure 3.2. Predicted values (marginal effects) of the interaction between Irony and Accent with mean f0 as a dependent variable.

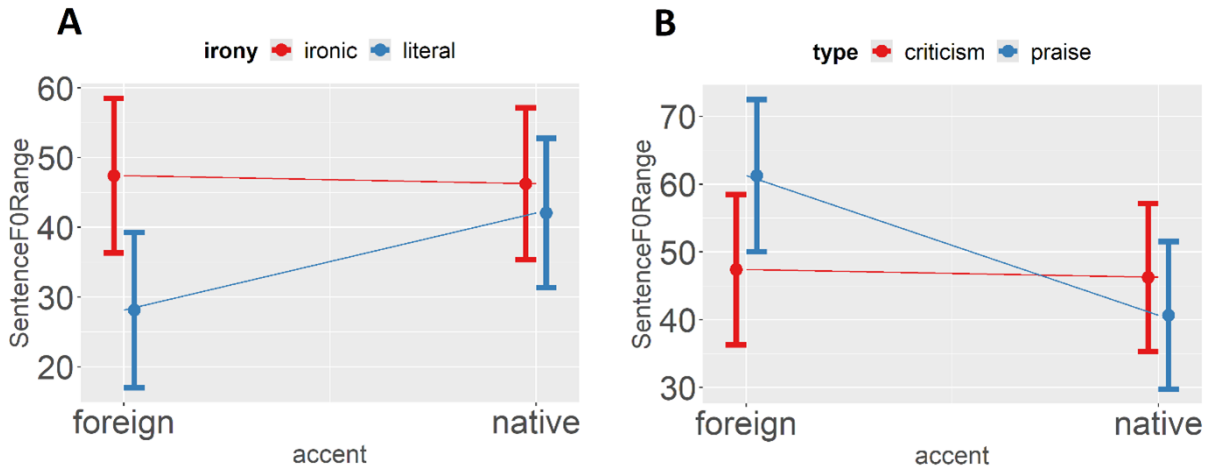

Figure 3.3. Predicted values (marginal effects) of the interaction between Irony and Accent (A) and Type and Accent (B) with f0 range as a dependent variable.

For mean intensity, there was a significant main effect of Irony only (t = -2.80, p = 0.006), with ironic utterances having a higher mean intensity than literal ones. This is in agreement with prior research (Anolli et al., 2000; Rockwell, 2000).

For intensity SD, there was a significant two-way interaction Irony × Accent ($t = 3.31$, $p = 0.001$) and a three-way interaction Irony × Type × Accent ($t = -2.12$, $p = 0.036$). Fig. 3.4 shows that foreign-accented speech was characterized by little difference in intensity SD between ironic praise and criticism ($p > 1$) but a much larger intensity SD for literal praise compared to literal criticism (t.ratio = 3.43 , $p = 0.009$). Both comparisons were insignificant for native speech (all ps > 1). However, there was a significant difference between literal and ironic criticism for native speech, with the latter having a smaller intensity SD (t.ratio = 3.67, $p = 0.004$), which is in agreement with prior research (Bryant, 2010; Bryant & Fox Tree, 2005).

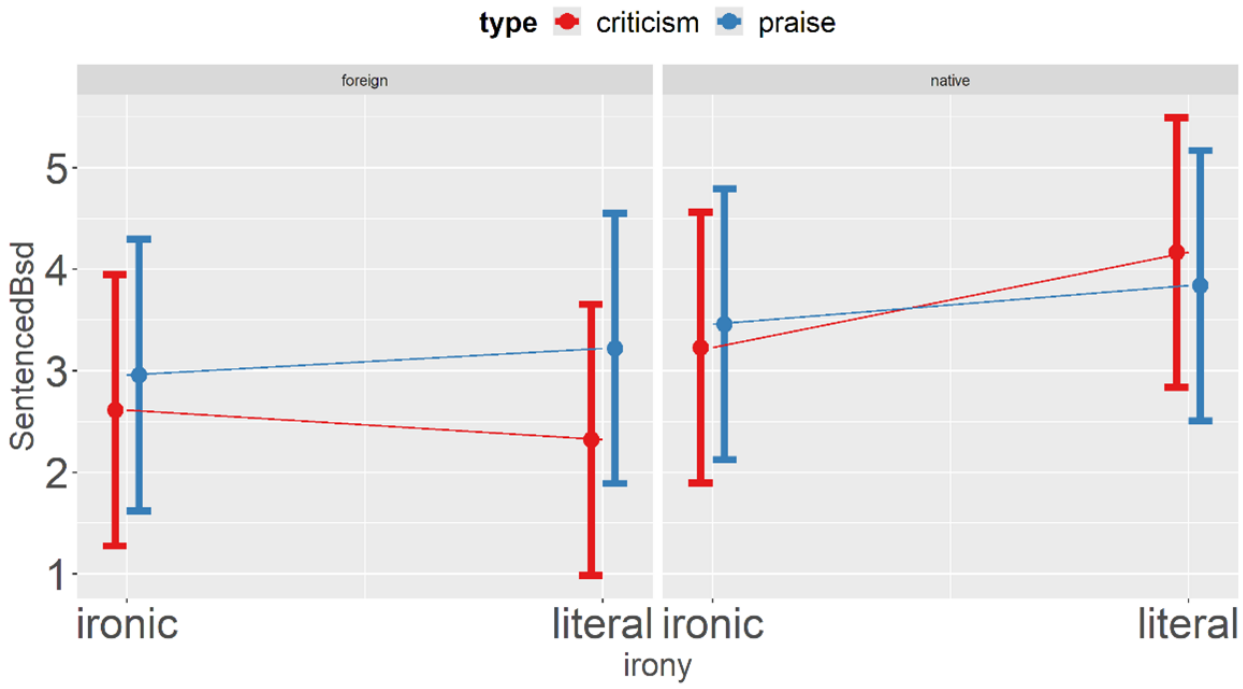

Figure 3.4. Predicted values (marginal effects) of the interaction between Irony, Accent, and Type and with intensity SD as a dependent variable.

For intensity range, there was a significant main effect of Irony ($t = -2.42$, $p = 0.017$) and a two-way interaction Type × Accent ($t = -6.62$, $p < 0.001$). While the difference between foreign-accented praise and criticism was insignificant ($p > 0.5$), native criticism had a much larger intensity range than praise (t.ratio = -7.94, $p < 0.001$) (Fig. 3.5). This is the first parameter in which the difference between native criticism and praise was so large and might be a unique marker of criticism, both literal and ironic, in English.

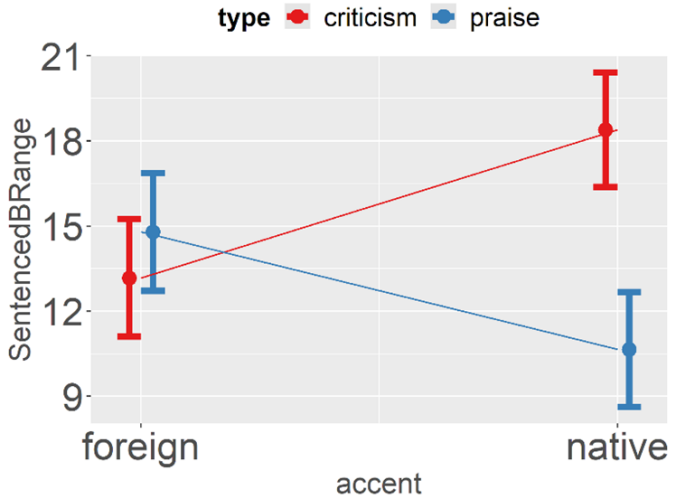

Figure 3.5. Predicted values (marginal effects) of the interaction between Irony and Type with intensity range as a dependent variable.

Summarizing, ironic utterances were characterized by a higher f0 SD and mean intensity regardless of accent and type. Importantly, accent interacted with either irony or type for virtually every measure. Foreign-accented utterances had a higher mean f0 and a larger f0 range in the ironic compared to literal condition, whereas almost no difference between the ironic and literal conditions was observed for native speech in this regard. Also, foreign-accented praise had a larger f0 range, larger intensity SD and intensity range than criticism. In contrast, native-accented criticism had a much larger intensity range than praise. This suggests that literal praise was more marked in foreign-accented speech and literal criticism was more marked in native speech. Most interesting, foreign-accented ironic praise was associated with a larger mean vowel duration than ironic criticism, whereas all native irony had longer mean vowel duration than literal speech regardless of its type.

EEG interval-level statistics. Table 3.2 and Fig. 3.6 provide the mean duration of EEG intervals by condition. For native speech, all ironic intervals were consistently longer than literal intervals. Ironic praise was slightly longer than criticism in every interval but the fifth, with the difference being the largest in the pre-target and target intervals. For foreign-accented speech, there was slightly more variability for ironic criticism and literal praise, even though the trend was essentially the same. The difference between ironic praise and criticism was more dramatic in non-native

speech, particularly in the pre-target and target intervals. In general, foreign-accented ironic praise was longer than any other condition in all intervals, which agrees with our prediction.

Table 3.2. Mean duration of EEG intervals (in ms) by condition. Standard deviations are in brackets.

Condition	EEG interval				
	#1	#2	#3	#4	#5
ironic criticism native	521(211)	475(127)	514(132)	375(110)	515(138)
ironic criticism foreign	585(254)	504(141)	556(98)	423(158)	509(117)
ironic praise native	524(188)	498(139)	534(124)	377(111)	511(128)
ironic praise foreign	601(232)	602(181)	605(106)	478(165)	515(148)
literal criticism native	457(187)	416(97)	478(106)	354(105)	478(100)
literal criticism foreign	498(235)	466(119)	528(97)	415(129)	463(115)
literal praise native	458(175)	392(98)	458(120)	345(107)	477(117)
literal praise foreign	514(243)	464(134)	565(146)	433(173)	464(123)

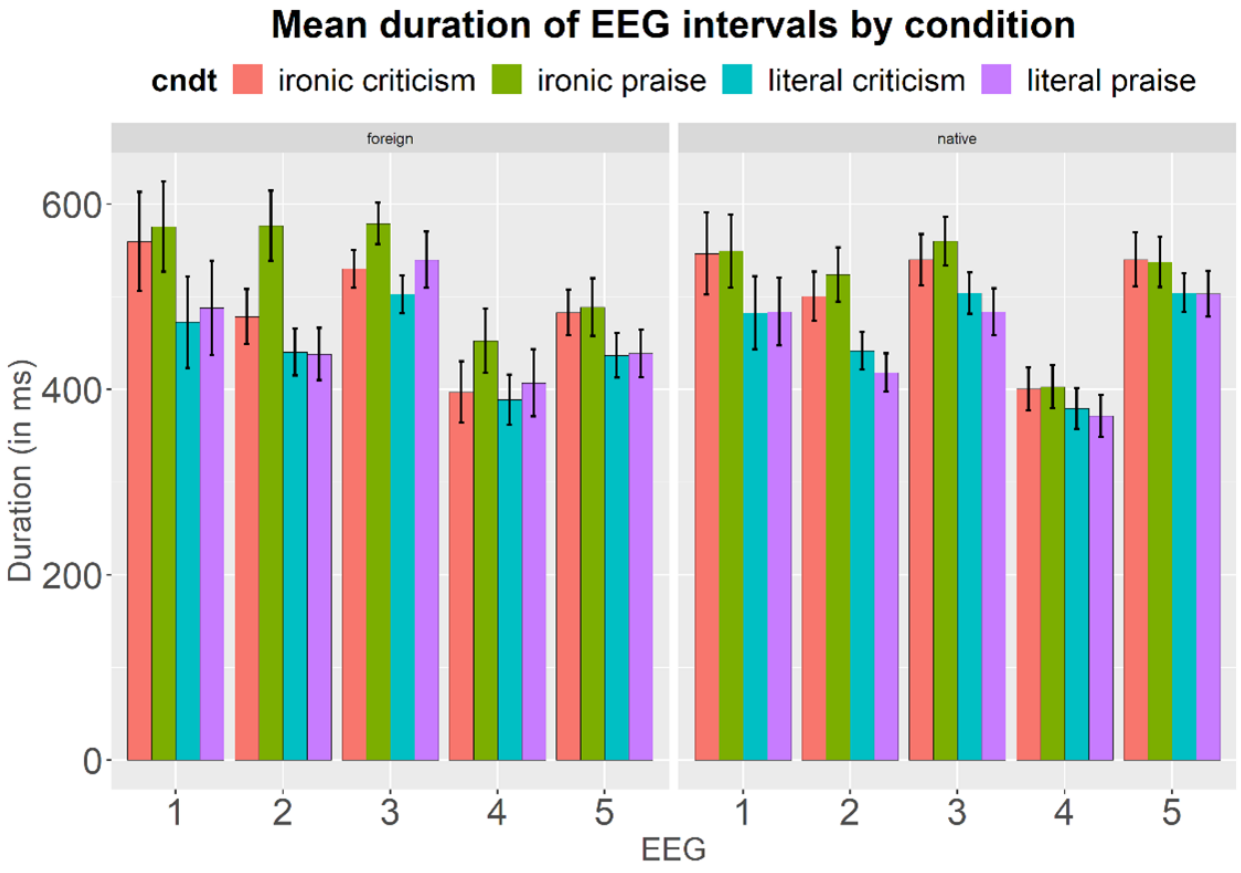

Figure 3.6. Mean duration of EEG intervals by condition. Black lines indicate error bars.

Linear mixed-effects modelling showed that Irony was the only significant predictor for the duration of EEG intervals one (t = -2.47, p = 0.014), three/target (t = -2.74, p = 0.007), and five/wrap-up (t = -2.39, p = 0.018), with literal intervals being shorter than ironic ones. For interval two/pre-target, there were main effects of Irony (t = -2.13, p = 0.034) and Type (t = 2.67, p = 0.008) and a significant interaction between Irony and Type (t = -2.28, p = 0.024). Post-hoc tests showed that ironic praise was significantly longer than literal praise (t.ratio = -5.36, p < 0.001) and that ironic praise was significantly longer than ironic criticism (t.ratio = 2.67, p = 0.034) (see Fig. 3.7). There was no difference between literal praise and criticism (p > 1). This suggests that ironic utterances were consistently longer than literal ones in most of the intervals, rather than just having longer targets. No effect of speaker's accent was found for any interval.

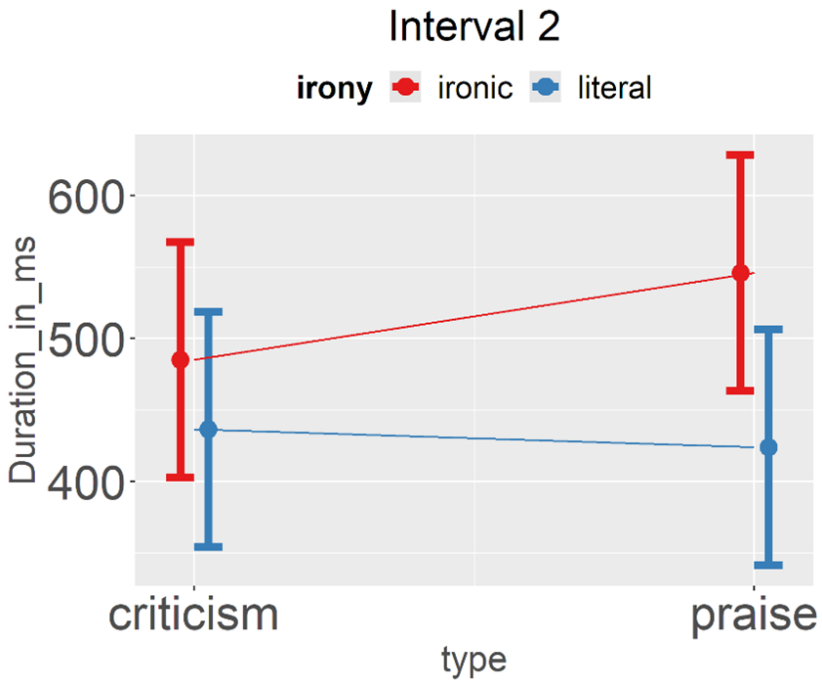

Figure 3.7. Predicted values (marginal effects) of the interaction between Irony and Type in interval two/pre-target with duration as a dependent variable.

Fig. 3.8 shows f0 at 10 equidistant points of the target interval. One of the most striking differences between accents was the variability of f0 by condition at point 1. Whereas all conditions look relatively the same in native speech, non-native speech exhibits high variation. Both literal conditions in foreign-accented speech start with a higher f0 than the ironic ones. Literal praise had the largest maximum f0 in non-native speech. In native speech, only literal criticism seemed

sufficiently different from the rest of the conditions, having much lower f0 at points four to seven. This is in line with the findings from utterance-level statistics that non-native speakers mark literal praise more than criticism, whereas native speakers mark literal criticism more than praise. Additionally, native and foreign-accented speech differed in the timing of the peaks for each condition. All foreign-accented conditions have peaks at very different points, whereas all the native ones have their peaks at point 5 or in-between points 5 and 6. Overall, foreign-accented speech had much more variance than native speech, which was expected. Fig. 3.9 shows f0 at 10 equidistant points for every EEG interval. One can see that the major differences between native and foreign accents are in intervals 4 and 5. In interval 4, the peaks for the two ironic conditions are at the very end of the interval in foreign-accented speech, whereas they are almost at the beginning of the interval in native speech. In the final interval, a much higher f0 was the key characteristic of literal praise in both accents, although the peak is located much later in foreign-accented utterances. Last but not least, f0 for ironic praise in interval one in foreign-accented speech was consistently higher at every time point except the first one than for any other condition.

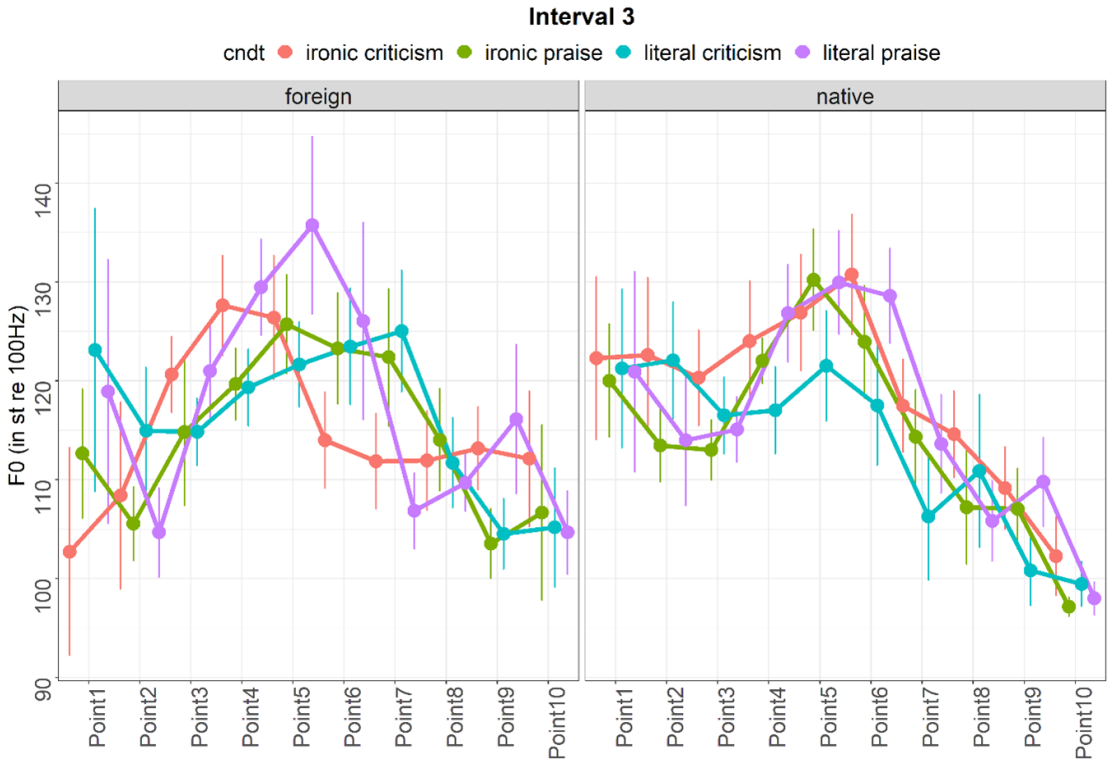

Figure 3.8. F0 at different time points of interval three/target by condition.

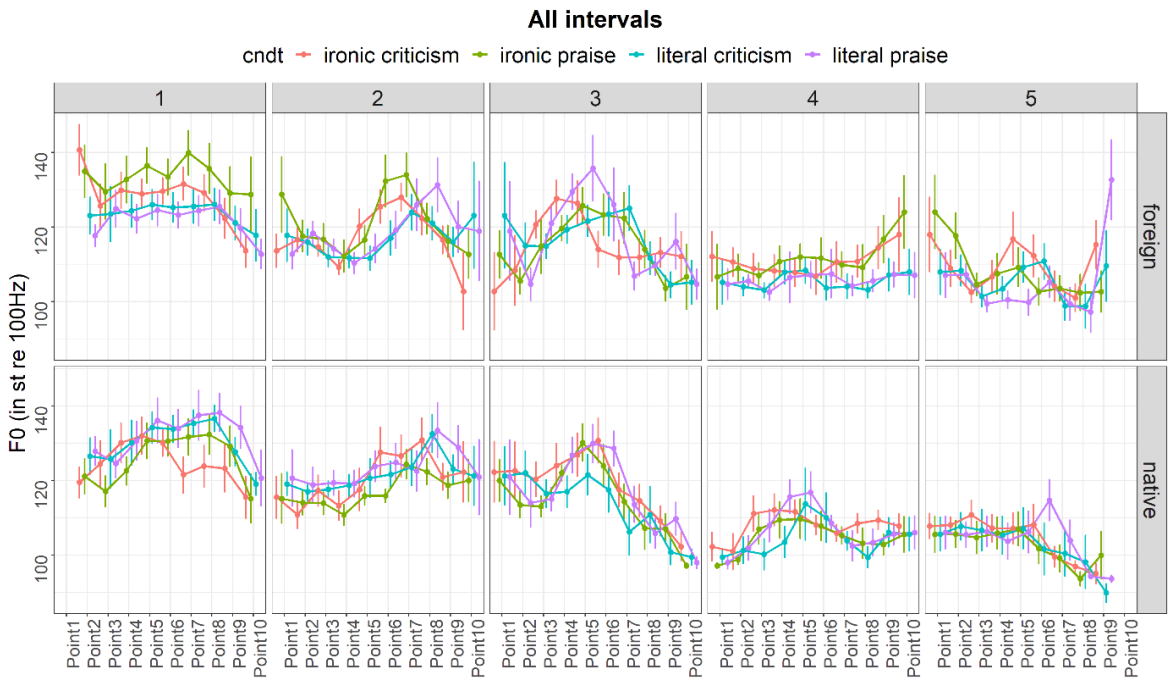

Figure 3.9. F0 at different time points of all EEG intervals by condition.

For maximum f0, linear-mixed effect modelling showed that Irony and the interaction between Irony and Accent were significant for intervals one (Irony: t = -5.01, p < 0.001; Irony × Accent: t = 4.20, p < 0.001), four/spillover (Irony: t = -3.41, p < 0.001; Irony × Accent: t = 2.06, p = 0.041), and five/wrap-up (Irony: t = -3.64, p < 0.001; Irony × Accent: t = 2.21, p = 0.028). No predictor was significant for intervals two and three (pre-target and target). Post-hoc tests showed that foreign-accented ironic speech had a larger max f0 than literal speech in interval #1 (t.ratio = -5.01, p < 0.001), #4 (t.ratio = -3.41, p = 0.003), and #5 (t.ratio = -3.64, p = 0.001), with no significant differences for native speech in any of those intervals (see Fig. 3.10).

We also ran an additional analysis by taking interval #3 as "target" and collapsing all intervals except #3 into "non-target". Thus, the starting model for backward selection had four predictors (irony, type, accent, and target). There was a significant main effect of Irony (t = -5.24; p < 0.001), two-way interactions Target × Irony (t = 3.15, p = 0.002) and Irony × Accent (t = 4.13, p < 0.001), and a three-way interaction Target × Irony × Accent (t = –2.47, p = 0.014). Foreign-accented speakers had the same maximum f0 between targets and non-targets in the ironic conditions (p > 1) but a much larger maximum f0 for targets than non-targets in the literal conditions (t.ratio = 4.38, p < 0.001), suggesting that they put a special emphasis on the target word in the literal

conditions only. In addition, foreign-accented literal non-targets had a much smaller max f0 than ironic non-targets (*t*.ratio = -5.24, $p < 0.001$), which suggests that all non-target segments were more marked in the ironic than in the literal conditions. In contrast, none of the comparisons were significant for native speech (all *p*s > 0.2) (see Fig. 3.11).

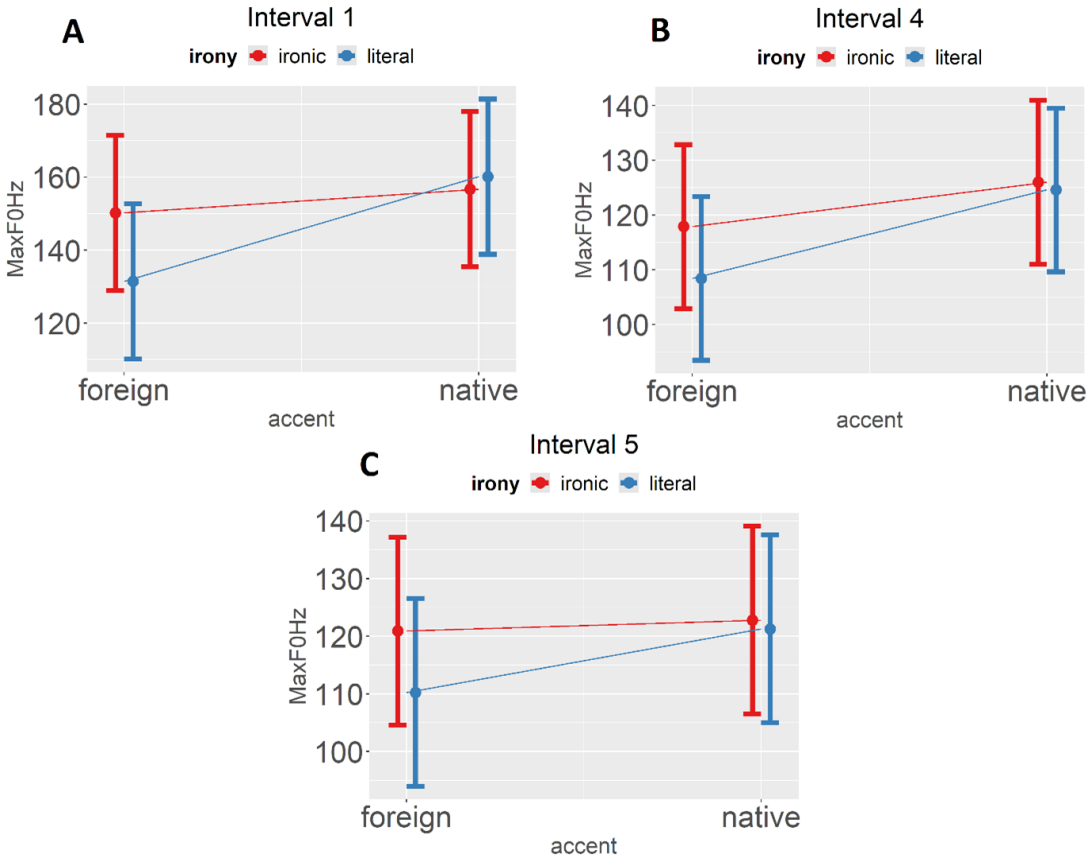

Figure 3.10. Predicted values (marginal effects) of the interaction between Irony and Accent in intervals one (A), four (B), and five (C) with maximum f0 as a dependent variable.

Fig. 3.12 shows mean intensity of EEG intervals by condition.

Accent was the only significant predictor for intervals one ($t = 5.36$, $p = 0.033$) and two/pre-target ($t = 5.84$, $p = 0.028$) and a marginally significant predictor for interval three/target ($t = 3.98$, $p = 0.058$), with native utterances being louder than foreign. For intervals four/spillover and five/wrap-up, the only significant predictor was Irony ($t = -2.21$, $p = 0.028$; $t = -3.18$, $p = 0.002$), with ironic utterances being louder than literal. In contrast to duration, this shows that the larger intensity of ironic utterances was likely due to the final part of the utterance, rather than being consistent across it.

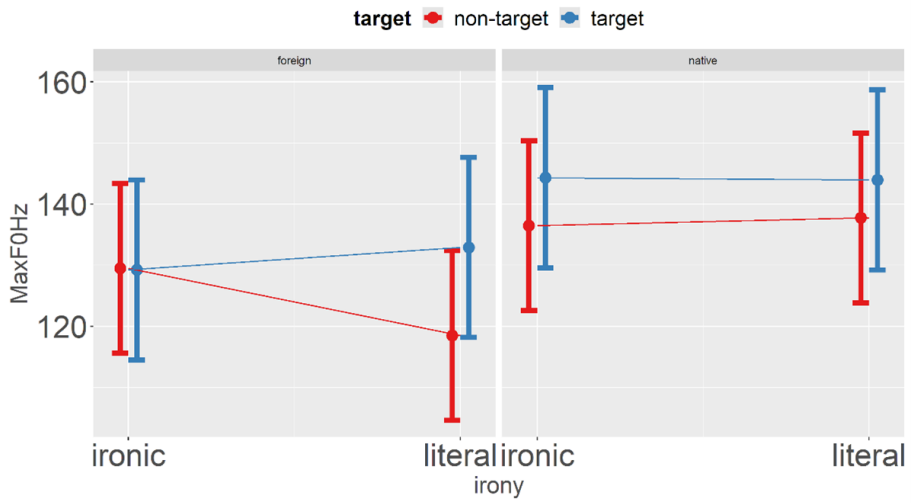

Figure 3.11. Predicted values (marginal effects) of the interaction between Irony, Accent, and Target with maximum f0 as a dependent variable.

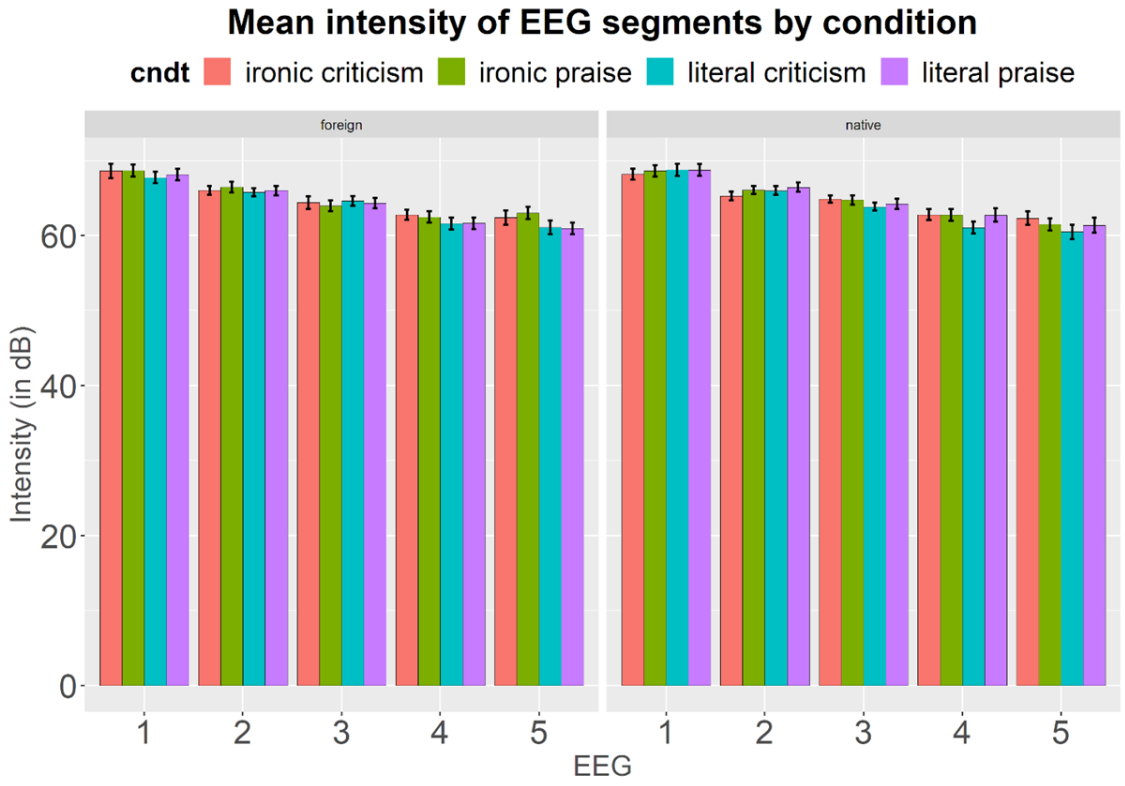

Figure 3.12. Mean intensity of EEG intervals by condition. Black lines indicate error bars.

We also ran a model with the Target predictor (target, non-target). For mean intensity, there was a significant main effect of Irony (t = -2.02, p = 0.045) and a marginally significant effect of Accent

($t = 3.61$, $p = 0.069$). Target and Type were not significant either by themselves or in interactions. Ironic utterances had a higher mean intensity than literal ones (this agrees with our utterance-level statistics reported above), and native-accented utterances had a marginally higher mean intensity than foreign-accented ones.

In summary, while some parameters turned out to be relatively consistent when analyzed at the level of EEG intervals (mean duration being larger in ironic than literal utterances), others did not (mean intensity). Also, interesting findings were observed for the Target predictor. Non-native speakers did not mark targets in the ironic utterances but marked them in the literal utterances. Also, non-targets were more marked (had larger max f0) in the ironic than in the literal utterances in foreign-accented speech.

Vowel-level statistics. Fig. 3.13 shows three acoustic parameters (mean duration, max f0, and mean intensity) for stressed vowels in the target interval by accent and condition.

Linear-mixed modelling revealed that no predictor was significant for either mean duration or max f0. For mean intensity, there was a marginally significant effect of Irony ($t = -1.83$, $p = 0.069$) and a significant interaction between Irony and Type ($t = 2.30$, $p = 0.023$). Stressed vowels in the ironic utterances had a larger mean intensity than in the literal ones. However, post-hoc tests did not show any significant contrasts in the Irony × Type interaction.

Finally, we also compared mean f0 and min f0 of stressed and unstressed vowels in target words in the ironic conditions, since an "inverse pitch obtrusion" was noted for ironic speech in the previous literature (Haiman, 1998, p. 31). We did not find evidence for the "inverse pitch obtrusion": stressed vowels had a significantly higher mean f0 ($t = 7.28$, $p < 0.001$) and min f0 ($t = 6.74$, $p < 0.001$) than unstressed ones.

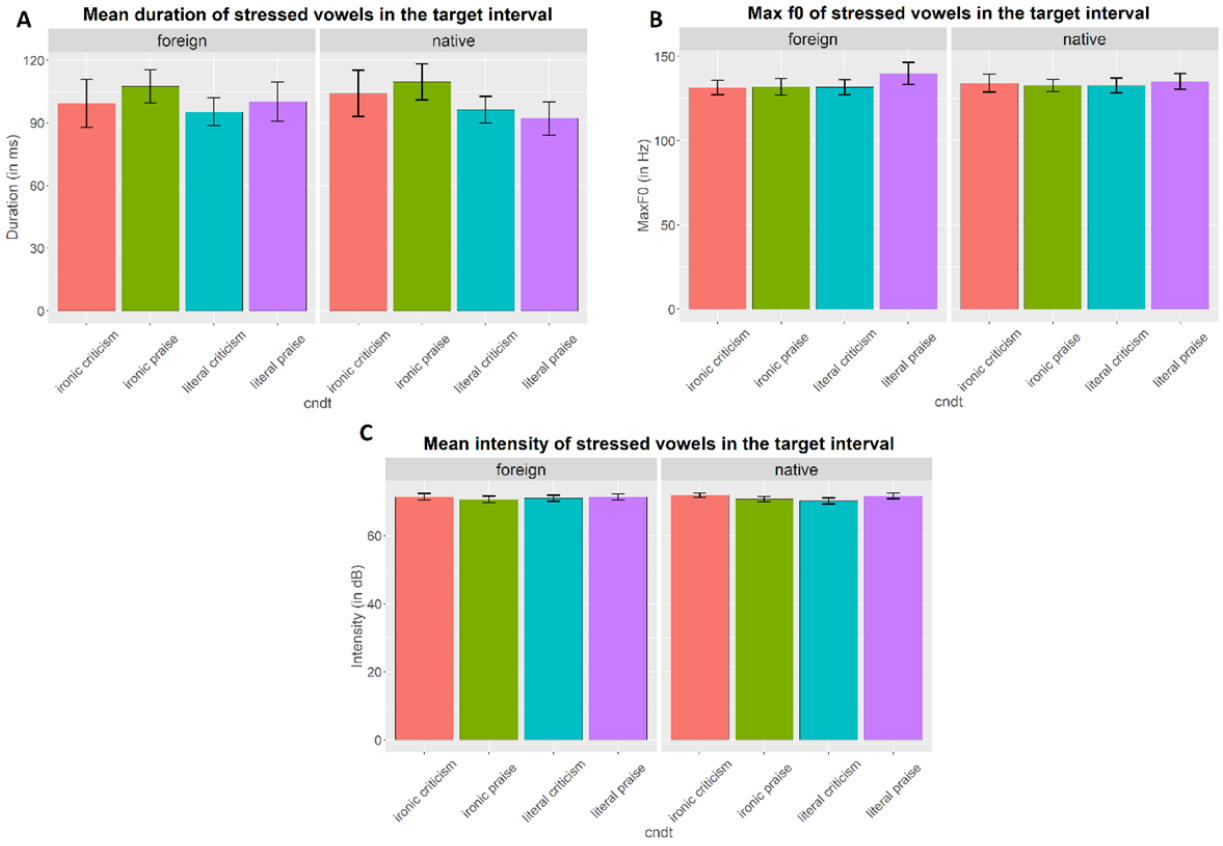

Figure 3.13. Mean duration (A), max f0 (B), and mean intensity (C) of stressed vowels in the target interval by condition. Black lines indicate error bars.

3.3 Discussion

The current study analyzed prosodic correlates of irony in native and Mandarin-accented speech. We examined two types of irony, ironic criticism (the most common one) and ironic praise (a much rarer one), comparing them to their literal counterparts. Most importantly, we analyzed the data at different levels of granularity (vowels, orthographic words, EEG intervals, and entire utterances). This allowed us to not only reveal more complex interactions with accent but also to test whether some prosodic contrasts at the utterance level found for irony in previous research and in our own study hold consistently across different parts of ironic utterances.

First, in agreement with some prior research (e.g., Bryant, 2010), ironic utterances were associated with a higher f0 variability, a larger f0 range, and a larger mean intensity than literal ones regardless of accent and type. In addition, a more detailed analysis showed that ironic utterances, in fact, had a larger mean intensity only in the post-target and wrap-up intervals (the final part of the utterance) rather than consistently across the utterance. This testifies to the importance of examining the same acoustic measures at various levels.

Second, in contrast to our coarse-grained analysis that did not find any interaction between irony and accent, the more detailed analysis revealed that accent, in fact, interacted with both irony and type in several measures and on several levels. We found that the difference in mean f0 and f0 range between ironic and literal conditions was much more pronounced in foreign-accented than in native speech. Mandarin-accented irony had a significantly higher mean f0, f0 range, and max f0 than literal speech. This only partially agrees with previous research on Mandarin which found higher mean f0 for ironic compliments in male speakers (S. Li & Gu, 2021) but lower f0 for ironic insults (S. Li et al., 2020). However, our study is not directly comparable to the above studies because they examined ironic insults and ironic compliments separately, comparing them to different baselines. Our findings could be explained by either linguistic transfer or, alternatively, by deliberate exaggeration of irony by non-native speakers. Because there are usually more communication breakdowns and misunderstanding between native and non-native speakers due to a lower language proficiency of the latter, it would be expected that they intonate and exaggerate ironic utterances more just to make sure that irony is sufficiently marked and easier to understand. We also found a much higher pitch variation in the target interval in non-native compared to native speech. All foreign-accented conditions had f0 peaks at very different time points of the target intervals, whereas all the native conditions had their f0 peaks roughly at the same point. This is more in line with the irony exaggeration theory because linguistic transfer is expected to be more consistent. Even more interestingly, foreign-accented speakers had roughly the same maximum f0 between targets and non-targets in the ironic conditions but a much lower maximum f0 for non-targets in the literal conditions. Foreign-accented ironic non-targets also had a significantly larger maximum f0 than literal non-targets. This also agrees a bit more with the exaggeration hypothesis.

One of the most interesting findings concerned mean vowel duration. In contrast to our previous analysis that found ironic utterances to be consistently longer than literal ones regardless of accent,

this measure interacted with accent. Native irony of both subtypes had a longer mean vowel duration than literal speech, in agreement with previous research on English (Anolli et al., 2000; Haiman, 1998). In contrast, foreign-accented ironic praise had a much longer mean vowel duration than ironic criticism (which, in turn, did not differ from literal conditions). This is in line with our predictions and suggests that non-native speakers may try to intonate ironic praise to the maximum to make sure it is not mistaken for actual criticism. Further research can compare ironic praise and ironic criticism in different accents in English to test the generalizability of our current findings for Mandarin-accented English. In addition, a direct comparison of the two irony subtypes in native Mandarin speech would be very helpful, since existing studies examined ironic criticism and praise separately. This would shed some light on whether longer vowel duration in ironic praise in Mandarin-accented English is a transfer from Mandarin or rather a more general phenomenon pertinent to foreign-accented speech for reasons outlined above.

In addition to our findings from a coarse-grained analysis that irony was associated with longer duration that literal speech, here we found that target intervals, as well as the beginning, the end of the utterance, and to some extent the pre-target interval, were consistently longer in the ironic than in the non-ironic utterances. This shows that, in contrast to mean intensity, mean duration seems to be more consistently and evenly longer in the ironic than in the literal utterances. Importantly, no interaction with accent was found for mean interval duration. Since previous research on Mandarin also reported slower duration for ironic compliments versus literal criticism (S. Li & Gu, 2021), this prosodic parameter may not differ between English and Mandarin (and, in fact, is consistently observed for all other examined languages, as discussed in the Introduction).

Surprisingly, we also found prosodic similarities between utterances of the same type (criticism or praise) in the absence of interaction with irony. For instance, foreign-accented praise (both literal and ironic) had a higher f0 range than criticism. Also, native criticism had a much larger intensity range than praise, while virtually no differences between the two were found for foreign-accented speech. These findings point toward the conclusion that criticism may be more marked than praise in English. Whether praise is more marked than criticism in Mandarin or just in Mandarin-accented English is not clear at this point and should be investigated by further research.

Naturally, our study had limitations. The main one is, of course, the low number of speakers per accent (two), which was due to COVID-related restrictions. Further research would need to

validate these findings with a larger number of speakers, potentially of different genders. In addition, since the stimuli were recorded as actual dialogs and the speakers constantly swapped roles, it remains possible that non-native speakers mimicked prosodic patterns of native speakers more than they would otherwise. We tried preventing this by counterbalancing the order in which the speakers started reading the lines, but some mimicking was probably inevitable. Future studies may opt for monologues instead of dialogues or dialogues between non-native speakers only to address this limitation.

Summarizing, irony in our study was associated with several prosodic markers, such as slower duration, higher f0 variability and a larger f0 range. We also found foreign-accented irony to be significantly more marked prosodically compared to native one with respect to almost all pitch-related measures. That said, the error bars in those measures, particularly in non-native speech, were large. Follow-up work could thus examine if such averages contain categorically different realizations (i.e. different pitch-accent types and/or phrasing). Moreover, Mandarin-accented ironic praise had significantly longer vowels than ironic criticism, suggesting that non-native speakers may invest more effort in prosodic marking of ironic praise to make sure it is successfully conveyed and not mistaken for literal criticism.

Chapter 4

Sorry, you make less sense to me: The effect of non-native speaker status on metaphor processing

Abstract

Preconceived assumptions about the speaker have been shown to strongly and automatically influence speech interpretation. This study contributes to previous research by investigating the impact of non-nativeness on perceived metaphor sensibility. To eliminate the effects of speech disfluency, we used exclusively written sentences but introduced their "authors" as having a strong native or non-native accent through a written vignette. The author's language proficiency was never mentioned. Metaphorical sentences featured familiar ("The pictures streamed through her head") and unfamiliar ("The textbooks snored on the desk") verbal metaphors and closely matched literal expressions from a pre-tested database. We also administered a battery of psychological tests to assess whether ratings could be predicted by individual differences. The results revealed that all sentences attributed to the non-native speaker were perceived as less sensical. Incorporating the identity of the non-native speaker also took more effort, as indicated by longer processing and evaluation times. Additionally, while a general bias against non-native speakers emerged even without oral speech, person-based factors played a significant role. Lower ratings of non-native compared to native speakers were largely driven by individuals from less linguistically diverse backgrounds and those with less cognitive reflection. Extraversion and political ideology also modulated ratings in a unique way. The study highlights the impact of preconceived notions about the speaker on sentence processing and the importance of taking interpersonal variation into account.

Keywords: linguistic expectations, non-native accents, metaphors, pragmatics, individual differences

4.1 Introduction

Imagine receiving a mysterious email with just one sentence: "Conscience is a burrowing mole"[10]. You cannot help but attribute some meaning to it. Metaphors, by nature, invite interpretation due to their semantic open-endedness (Boyd, 1993; Gibbs et al., 1991). Now, suppose you are told beforehand that the metaphor is created by either a renowned poet or a computer program with a random generator. Would that influence how meaningful it appears to you? Research shows that the answer is yes–you would rate the poet's sentence as more meaningful than the computer-generated one (Gibbs et al., 1991). Thus, the knowledge you have about the author immediately influences how you perceive the sentence.

This, of course, is not an isolated example of how speaker-induced expectations shape language processing. Traditionally, linguistic accounts posited a clear dichotomy between the context-invariant "semantic" meaning derived from the lexical meaning of words, and the extended "pragmatic" meaning refined by the available extralinguistic information, such as the communicative context in which the conversation takes place or the speaker's identity (Grice, 1975; Levinson, 2000). However, this strict separation has been challenged by recent findings, particularly by studies using time-sensitive methodologies (Bornkessel-Schlesewsky et al., 2013; Hagoort et al., 2004; Regel et al., 2011; Van Berkum et al., 2009b; Van Berkum, van den Brink, et al., 2008). Although the debate on the exact relationship between semantics and pragmatics remains ongoing, there is growing consensus that assumptions about the speaker are rapidly and automatically taken into account during language comprehension. For instance, listeners exhibit cognitive difficulty when the perceived speaker's identity contradicts societal stereotypes (Pélissier & Ferragne, 2022; Van Berkum, van den Brink, et al., 2008). They also adjust their pragmatic inferences based on how knowledgeable they think the speaker is (Bergen & Grodner, 2012), what they know about the speaker's communicative style (Regel et al., 2010), and whether they believe the speaker is able to act on their statement (Bornkessel-Schlesewsky et al., 2013).

Linguistic identity, specifically the person's accent, plays a major role in shaping listener expectations. Broadly speaking, accent is a person's unique manner of pronunciation (Giles, 1970)

[10] Whiting (1942, p. 870), as cited in Katz et al. (1988).

and thus serves as an important social cue. A non-local L1 accent may reveal that this native speaker comes from a different region of the country, whereas an L2 accent commonly indicates a foreigner who immigrated into a country after puberty (Dollmann et al., 2020; Piske et al., 2001; Scovel, 2000; Tahta et al., 1981). Thus, even though accent is not directly inferable from a person's appearance, especially in multicultural societies like Canada or the United States, it forms an integral part of a person's identity and conveys important information about their geographical and ethnic background, as well as social class (Pélissier & Ferragne, 2022). Further, L2 accent is a multifaceted phenomenon that affects communication in at least three different ways: through disfluency, through transmitting a signal of incomplete L2 mastery, and through transmitting a signal of cultural foreignness. We discuss these three effects in detail below.

First and foremost, non-native speech acoustically deviates from native speech on levels from purely phonetic to prosodic (Gut, 2012; Hanulíková & Weber, 2012). For instance, non-native speakers often replace L2 phonemes that do not exist in their L1 with a native sound (Wester et al., 2007). Listening to foreign-accented speech has been compared to listening under adverse conditions, since it shares many characteristics with acoustically degraded speech (Van Engen & Peelle, 2014). Pairing foreign accents with the speaker's face facilitates processing but still incurs greater processing cost than native accents (Grey et al., 2020). Since it has been shown that adverse listening conditions disrupt processing fluency (H. Lane, 1963; Munro, 1998) and lead to harsher evaluations of the speaker (Dragojevic & Giles, 2016), it is not surprising that foreign accent strength correlates with more negative speaker attitudes (Dragojevic et al., 2017). Non-native speakers have time and again been shown to be judged harsher–less reliable, less intelligent, less successful (Fraser & Kelly, 2012; Fuertes et al., 2012) and possibly even less credible (Lev-Ari & Keysar, 2010; cf. Souza & Markman, 2013) just because of their accent.

Second, due to the still developing L2 command, the language of non-native speakers tends to be error-prone and in general less reliable in conveying their intended message. As an important consequence, because listeners assume higher error rates in foreign-accented speech, they may adjust their manner of processing accordingly, engaging in more shallow (Lev-Ari & Keysar, 2012) and less literal (Gibson et al., 2017) processing strategies. For instance, Gibson and colleagues found that implausible sentences such as "The businessman benefited the tax law" were more likely to be interpreted in a plausible way when delivered with a foreign accent. Additionally,

people may adopt a more lenient attitude in their linguistic judgments and give non-native speakers the benefit of the doubt. Research by Fairchild et al. (2020) and Fairchild & Papafragou (2018) demonstrated that pragmatically infelicitous sentences such as "Some people have lungs" are rated as more sensical when attributed to non-natives speakers. On the flip side, since non-native speakers are perceived as grammatically unstable, listeners are not only less surprised by their mistakes (Hanulíková et al., 2012) but may also perceive even grammatically correct sentences as less grammatical when they are foreign-accented (Wesołek et al., 2023).

And lastly, since non-native accent typically signals that the speaker is a foreigner, monolingual native listeners may instantly categorize them as an outgroup (Ryan, 1983). This entails reduced common ground and, unsurprisingly, has been found to affect language processing, particularly figurative language. Generally speaking, common ground between interlocutors plays a critical role in both production and comprehension of non-literal language. When cultural knowledge, experiences, and attitudes are not shared, listeners experience significantly more trouble with identifying ironic intent (Kreuz & Link, 2002), and speakers tend to use less irony overall (Averbeck, 2015; Averbeck & Hample, 2008). Since native and non-native speakers are presumed to have less common ground, it is non-surprising that irony has been found to be perceived as less ironic when produced by non-native speakers (Bazzi et al., 2022; Caffarra et al., 2018; Puhacheuskaya & Järvikivi, 2022) and cause more processing difficulty on a neural level (Caffarra et al., 2019). Native listeners are also less confident when interpreting foreign-accented irony (Puhacheuskaya & Järvikivi, 2022).

Despite all the above, the intersection of non-native speaker identity and metaphor processing has not been explored yet. Metaphors, irony, rhetorical questions, indirect requests, jokes, and other types of figurative language requiring inferences all have an advantage over ungrammatical sentences in that both native and non-native speakers can and do use them regularly. Some evidence suggests that we use non-literal language on a daily basis (Gibbs, 2000; Hancock, 2004; Whalen et al., 2009), which makes it very well-suited for exploring the effect of non-nativeness. As discussed in the beginning of the article, the information the reader possesses about the author of a particular metaphor can immediately constrain and shape its interpretation (Gibbs et al., 1991). It is thus reasonable to hypothesise that information about the speaker's non-native status, since it usually correlates with less intentionality in semantic choices, less reliability in conveying the

message, and higher error rate due to lower language proficiency, will also immediately shape how meaningful the metaphors appear–and, perhaps, even literal language. Although there is some evidence that comprehending metaphors is more cognitively demanding than comprehending literal sentences (Arzouan et al., 2007; Bambini et al., 2016; Lai et al., 2009; Obert et al., 2018), and that deriving meaning from literal versus metaphorical sentences recruits distinct neural mechanisms (Stringaris et al., 2007), Gibbs et al. (1991) found that the readers also rated *literal* sentences as less meaningful when they were attributed to a computer program.

Importantly, large interpersonal variation has been attested in figurative language processing and pragmatic inferences in general (Abraham et al., 2021; Olkoniemi et al., 2016; Puhacheuskaya & Järvikivi, 2022; Stamenković et al., 2019; Werkmann Horvat et al., 2022). Most research on individual differences has historically focused on the Autism Quotient (AQ) scores and working memory measures, although this is slowly changing. For instance, Mayn & Demberg (2022) examined a wide range of individual differences in a pragmatic reference game and found that the Cognitive Reflection Test scores (Frederick, 2005) significantly predicted the ability to draw pragmatic inferences (but not the AQ scores or working memory). Puhacheuskaya & Järvikivi (2022) additionally showed that participants' political leaning predicted irony comprehension accuracy, so that higher conservatism scores were correlated with poorer irony detection. For metaphor processing specifically, individual differences in creativity, working memory capacity and need for cognition (NFC) (Cacioppo & Petty, 1982), non-verbal intelligence (as measured by the nonverbal Raven's Progressive Matrices test; Arthur & Day, 1994) and multilingualism have all been found to affect metaphor comprehension (Abraham et al., 2021; Olkoniemi et al., 2016; Stamenković et al., 2019; Werkmann Horvat et al., 2022). Moreover, since our goal is to examine whether metaphor processing is affected by the non-native speaker status, it is also important to consider individual variation in non-native speech processing and interpretation. For instance, Dewaele & McCloskey (2015) showed that childhood exposure to different ethnicities and multilingualism, the diversity of one's workplace, speaking multiple languages, and, surprisingly, being older were correlated with being less bothered by foreign accents. These more laid-back attitudes theoretically can facilitate pragmatic inferences or sensibility judgments. However, Fairchild & Papafragou (2018) examined cultural attitudes toward Chinese-American speakers and found that they did not correlate with ratings of pragmatically infelicitous sentences for either

native or non-native speakers. More research is thus needed. Overall, recent psycholinguistic research has shown that both online and offline linguistic measures are affected by individual differences, from empathy to disgust sensitivity to political leaning (Eekhof et al., 2021; Hammond-Thrasher & Järvikivi, 2023; Hubert Lyall, 2019; Hubert Lyall & Järvikivi, 2021; Van Berkum et al., 2009b).

Thus, in addition to the effects induced by the non-native speaker status, we decided to do an exploratory investigation of possible character traits that may mediate those effects. Specific hypotheses for the study as well as a full battery of tests and tentative predictions for these measures are outlined in the next section.

4.1.1 The present study

The main objective of this study was to investigate whether people perceive metaphoric sentences differently if they are led to believe that these sentences were created by non-native as compared to native speakers. As a dependent measure we chose sensibility ratings since it is a very straightforward task for the participants and does not require metalinguistic judgments about metaphors.

Before we discuss our hypotheses for the study, it is important to address a methodological question. While expectations related to the still developing L2 command and social group membership generally cannot be removed from the equation, in a sense that they are inherent in non-native speech processing, negative affect due to disfluency can be manipulated. Naturally, most previous research recorded sentences in a native versus foreign accent and then tested the participants using auditory stimuli, which makes it near-impossible to disentangle the effect of processing difficulty due to reduced intelligibility from the other two effects. Although there is no perfect solution to this, some studies employ auditory stimuli read by a native speaker but devise a cover story that the stimuli were originally created by a non-native speaker instead (Foucart & Hartsuiker, 2021). Another method that is somewhat gaining popularity is to use no speech at all and instead introduce sentence “authors” through a written vignette presented either before the experiment or before each block of stimuli (Bazzi et al., 2022; Fairchild et al., 2020; Foucart et al., 2019). Both of these methods, while being riskier in terms of authenticity and believability, have

an advantage of controlling for any negative affect from extra cognitive load. That is why we opted for exclusively written materials in this study.

Since metaphor comprehension has not been investigated in this particular context before, even more so when separated from actual oral speech, formulating specific predictions was tricky. Taking the prior research into account, we had two rival hypotheses for the study.

The leniency hypothesis (H1): The participants may rate metaphoric and possibly literal sentences as *more* sensical when they are attributed to non-native speakers. Namely, since people assume a higher incidence of errors in non-native speech and less control over semantic choices, they may exhibit greater willingness to re-interpret non-native sentences for a more plausible version, interpret them in a less literal fashion, and show greater leniency and charitability when making sensibility judgments (Fairchild & Papafragou, 2018; Gibson et al., 2017). In short, the process of deriving meaning from non-native metaphors may be facilitated.

The downgrade hypothesis (H2): The participants may rate metaphoric and possibly literal sentences as *less* sensical when they are attributed to non-native speakers. Namely, again, since people generally assume lower language proficiency, less agency in semantic choices due to more limited vocabulary, and a higher incidence of errors, non-native sentences may appear less meaningful, similar to how sentences generated by a computer program appear less meaningful (Gibbs et al., 1991). Metaphorical sentences in particular may be reanalyzed as a poor choice of word for the context, awkward phrasing, a semantic mistake, or even a possible mistranslation from the non-native speaker's L1, hindering interpretation.

We also explored the effect of metaphor familiarity. It has been widely attested that familiar and unfamiliar/novel metaphors are processed differently (Arzouan et al., 2007; Blasko & Connine, 1993; Cardillo et al., 2012; Tartter et al., 2002). Previous research using native-accented or written stimuli suggested that novel metaphors seem to initially appear anomalous, are much harder to process, and it takes much longer to derive a figurative meaning from them. The graded salience hypothesis (Giora, 1999) provides a theoretical basis for these effects. According to this account, the salient meaning (i.e. conventional, frequent, familiar, enhanced by prior context) is always computed first and the non-salient one is computed second. This would mean, for instance, that the metaphoric meaning of only novel, but not conventional/frequent/familiar metaphors, is accessed after the literal meaning has been processed and rejected. How exactly this would affect

sensibility ratings when novel metaphors are attributed to non-native speakers is hard to predict. ERP studies on foreign-accented semantic anomalies suggest that meaning repair as evidenced by the P600 may not happen when listening to foreign-accented speech and that semantic anomalies might be treated as straight-out errors (Romero-Rivas et al., 2015). Perhaps one could thus hypothesise that novel metaphors may be more easily treated as semantic errors when they are attributed to non-native speakers and should thus exhibit the lowest sensibility ratings of all the conditions.

Additionally, we examined the following range of individual differences: the participants' language background to estimate their exposure to different accents, their explicit (via a new survey) and implicit (via a modified IAT test; Greenwald et al., 1998) language attitudes and biases against foreigners, Big-5 personality traits (John & Srivastava, 1991), political views (Wilson & Patterson, 1968), and the Cognitive Reflection trait (Frederick, 2005). Since examining individual variation in psycholinguistic research is still in its very infancy, this portion of the study was largely exploratory. Due to the limited research on how character traits may affect language processing, we cannot formulate solid predictions based on previous studies. Nevertheless, we describe below our tentative predictions based on existing literature and theoretical frameworks.

For language background, we predicted that participants with less accent exposure would be less lenient toward non-native speakers than those with more exposure (Dewaele & McCloskey, 2015). In addition, we expected to see an effect of bilingualism interacting with both metaphor conditions (Werkmann Horvat et al., 2022) and speaker conditions due to bilinguals usually growing up in more diverse environments and being potentially less likely to categorize non-native speakers as social outgroups than monolinguals (Ryan, 1983) which may foster a more positive attitude. For political views, since they robustly correlate with anti-immigrant prejudices (Banton et al., 2020; Hodson & Dhont, 2015) even despite exposure and immigrant friends (Kiehne & Ayón, 2016), it might be that more conservative participants will rate metaphoric sentences by non-native speakers as less sensical because of more negative affect toward them and because they might be less likely to invest effort in deriving a figurative meaning. For the same reason, we predicted that people with more negative explicit and implicit accent attitudes would be affected in the same way. Puhacheuskaya & Järvikivi (2022) also found that more right-leaning participants showed lower irony detection accuracy, and this effect might be generalizable to metaphoric sentences. For

cognitive reflection scores, since Mayn & Demberg (2022) found them to be significantly correlated with the ability to draw pragmatic inferences, we predicted that people with higher scores might rate native and, perhaps, even more so non-native, metaphors as more sensical. In general, this measure reflects the ability to override initial, intuitive responses and come up with a better answer, so it may be particularly predictive of unfamiliar metaphor ratings. And for Big-5 personality traits, we hypothesized that Openness, because it indexes willingness to engage with and appreciation for novel ideas, and Extraversion, because it indexes wide social circles, pleasure from human interaction and emotional expressiveness, might predict higher ratings of metaphorical sentences in general and metaphorical sentences attributed to non-native speakers in particular.

4.2 Methods

4.2.1 Participants

A total of 98 self-identified native English speakers were recruited through Prolific. They received the equivalent of £3.75 in USD for their participation as per Prolific guidelines. The participants were recruited according to quotas (gender: 50% male, 50% female; political orientation: 50% liberal, 50% conservative). Data from 13 participants were excluded (N = 3 due to two or more failed attention checks out of four according to Prolific guidelines and N = 10 due to total comprehension accuracy below 70%). The final sample thus included 85 participants (mean age: 45, range: 20–77, SD: 16). The gender composition was as follows: 42 women (49%), 41 men, 1 non-binary, and 1 trans. Highest degree obtained was as follows: 25 completed high school, 14 completed technical/vocational school, 30 had Bachelor's degree, 12 had Master's degree or higher, 1 preferred not to respond and 3 selected "other". All self-identified as American.

4.2.2 Materials

We used a pre-tested metaphor database by Cardillo et al. (2010) that contains pairs of metaphorical and literal sentences closely matched along multiple dimensions, including familiarity, naturalness, and imageability. The database has both nominal and verbal metaphors, which are further subdivided into auditory (e.g., "The headline buzzed in his ears") and motion

(e.g., "The detective jumped at the clue") types. We decided to use verbal metaphors only because they blend well with the sentences we chose as fillers (see next paragraph) and because they are a little less obvious as metaphors than nominal ones (of the type "X is Y"). In addition, an fMRI study using verbal and nominal metaphors from the same database did not find any neural differences between the processing of the two types (Cardillo et al., 2012). We divided the auditory and motion portions of the verbal metaphor database into two halves using the median value for metaphor familiarity. Eight metaphors from each part were selected to create two conditions: familiar and unfamiliar. Average lexical frequency[11] of all items in the metaphoric sentences was kept below 190 per million (mean = 47, median = 22), lexical frequency of the verb below 22 per million (mean = 2, median = 0.24). All characteristics of the metaphorical sentences are provided in Table 4.1. According to Student *t*-tests, valence, verb frequency and average lexical frequency did not differ between metaphor conditions (ts < 1.4, ps > 0.1), whereas familiarity, figurativeness, naturalness, and imageability did (ts > 4, ps < 0.001).

Table 4.1. Lexical characteristics of the metaphorical sentences used in the study: Means and standard deviations (in parentheses).

metaphor condition	valence	figurativeness	familiarity	naturalness	imageability	mean freq of content words	verb freq
familiar	0.29 (0.3)	5.5 (0.6)	5.0 (1.0)	5.0 (0.9)	3.9 (1.2)	53.3 (42.8)	3.6 (6.2)
unfamiliar	0.16 (0.2)	6.4 (0.3)	2.0 (0.5)	2.2 (0.5)	2.4 (0.5)	40.8 (55.7)	0.8 (1.1)

As fillers, we used sentences with and without violations of gender stereotypes from Osterhout et al. (1997). These sentences were chosen because they agree well with the sensibility rating task used in this experiment, can be plausibly attributed to both native and non-native speakers, and are similar in structure and length to our main stimuli (e.g., "The secretary bought himself a plane ticket."). Fillers were of four different types: female/male stereotype violation/match.

In total, the stimuli file contained 32 metaphors, 32 literal counterparts, and 32 fillers. Examples of experimental sentences with critical verbs italicized are provided below:

1. The newspaper *pounced* on the story. / The lion *pounced* on its prey. (metaphor familiar / literal)

2. The flowers *purred* in the sunlight. / The kitten *purred* on the sofa. (metaphor unfamiliar / literal)

[11] Both the average lexical frequency of all items and lexical frequency of the verb come from Brysbaert & New (2009).

In total, the stimuli file contained 32 metaphors, 32 literal counterparts, and 32 fillers. Half of the metaphors were auditory, and half were motion. The equal number of motion and auditory metaphors served purely as a control measure to ensure balance across subtypes, and we did not have any predictions regarding these categories. Of each of those types, half were familiar and half were unfamiliar. The list of stimuli is available in Appendix A.2. The full table with the stimuli and their properties is available on OSF.

4.2.3 Procedure

The experiment was programmed in PsychoPy3 and run via Pavlovia (Peirce et al., 2019). It consisted of two blocks, a native speaker and a non-native speaker one (the order counterbalanced between participants). The following small cover story was created to introduce the task: "For this experiment, we asked native and non-native speakers of English in one of our earlier studies to come up with 50 short sentences (we will call them 'sentence authors'). Your task in this experiment will be to judge how much sense their sentences make." Before each block, the participant was presented with a written vignette of a person whose statements they had to rate afterwards. Each vignette came in versions A and B, which differed only in the alleged university major of the student and their hobbies. The vignettes were modified from Fairchild & Papafragou (2018) and are provided below:

Native Speaker: Emma is an undergraduate student at the University of Washington, majoring in History/Sociology. She is doing well in her classes and plans to be a high school teacher after graduation. Emma moved with her family to Seattle from Alberta. Emma has such a strong Canadian accent that her classmates often make fun of her. In her spare time, Emma likes to hike/run and play the piano/guitar.

Non-native Speaker: Zhou is an undergraduate student at the University of Washington, majoring in History/Sociology. She is doing well in her classes and plans to be a high school teacher after graduation. Zhou moved with her family to Seattle from China. Zhou has such a strong Chinese accent that her classmates often make fun of her. In her spare time, Zhou likes to hike/run and play the piano/guitar.

The native speaker was chosen to be from Canada[12] so as to make them a partial ingroup/outgroup (an immigrant from a different country but a linguistic ingroup due to being a native English speaker) while the non-native speaker was a double outgroup (an immigrant from a different country and not a native speaker). It also made it more natural for a native speaker to "have a strong accent" when they allegedly come from a different country. Together with the bio we showed a photo of a mixed-race woman who most participants in the norming study identified as Asian. The photos were taken from the Chicago Face Database (Ma et al., 2015). Photo 1 was rated as follows: age 22.4; Asian Probability 0.57, Latino 0.17, multi 0.25, attractive 4.03. Photo 2 was rated as follows: age 26.7; Asian Probability 0.69; Latino 0.24, multi 0.07, attractive 4.41. This ensured that both pictures represented people of similar age and attractiveness rating and mixed-race people who were still mostly rated as Asian. The photos were counterbalanced between participants.

After the bio, three comprehension questions followed in a random order to ensure that the participants read the vignette (Where is X from? What is X majoring in? What does X like to do in her spare time?). The participants were then asked to rate sentences that followed on how much sense they make (no sense = 0 to perfect sense = 50). Same as in Fairchild & Papafragou (2018), the participants were told that they should make use of intermediate values on the scale because sentences can make more or less sense. The participants' response times were also recorded.

Since metaphorical sentences and literal sentences lexically overlapped in only the verb (e.g., "The headline buzzed in his ears" vs. "The bees buzzed in the garden"), all stimuli could be used in a within-participant design, ensuring more control. A Python script was used to randomly generate nine experimental lists that satisfied the following criteria: 1) if a verb appeared as a metaphor in a native block, its literal form appeared in a non-native block; and 2) both native and non-native blocks contained the same number of items in each value of Metaphor/Literal/Filler conditions and all their subtypes (i.e. 4 sentences in the "metaphor auditory familiar", 4 sentences in the "metaphor auditory unfamiliar", etc.). Overall, there were 16 metaphors per each block. Each block was also

[12] One could argue, of course, that a Canadian accent in English is, in a sense, foreign, since it is an accent that comes from a foreign country when the reader themselves is from the U.S. However, what is important is that the Canadian speaker in our study "created" sentences in their native language, whereas the Chinese speaker in their foreign language, hence the native/non-native distinction.

pseudorandomized so that no more than three repetitions of the Metaphor/Literal/Filler condition could occur in a row. Fig. 4.1 shows an example of a trial from a native and a non-native block.

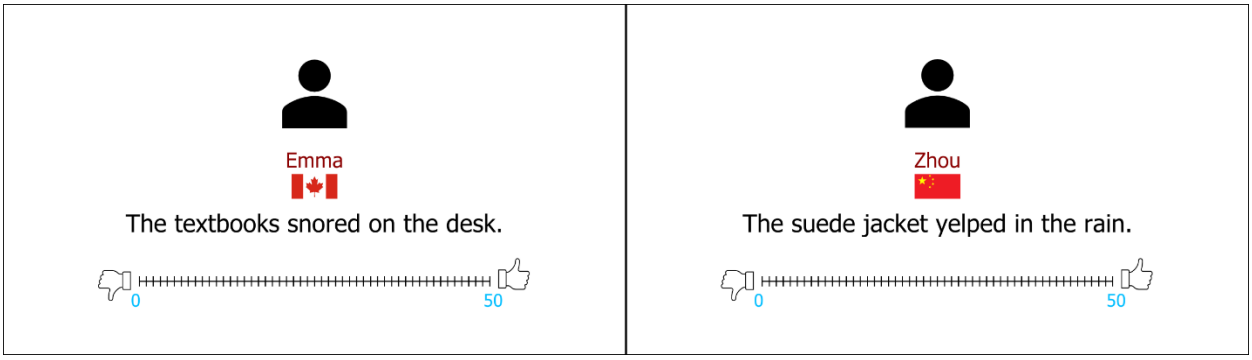

Figure 4.1. An example of a trial with an unfamiliar metaphor from a native (on the left) and non-native (on the right) block.

After the participants finished rating 96 sentences (48 per speaker, with a break in-between) they were asked to complete the survey portion. The description of each survey is provided below.

Language Background Questionnaire. We collected detailed background information about each participant regarding the linguistic diversity of their childhood and current environment, such as the number of parents/relatives/friends whose first language is not English, the number of foreign languages they speak, whether they are bilingual, etc. The full list of questions is available in Appendix B.4.

Wilson-Patterson Conservatism (W-P). We measured participants' political ideology along the left-right dimension using the 20-item Wilson-Patterson Conservatism Scale (W-P) (Wilson & Patterson, 1968). The full list of questions is available in Appendix B.1.

Big-Five Inventory (BFI). We used the complete 44-item version of the Big-Five Inventory (John & Srivastava, 1991).

Cognitive Reflection Test (CRT). Similar to Mayn & Demberg (2022), we used a 10-item version of the Cognitive Reflection Test that contained 3 math questions, 3 verbal questions, and 4 "decoy" questions. The order of the items was randomized. Since familiarity with the questions may affect scores on the test (Stieger & Reips, 2016), we also ran a survey in which the participants had to indicate whether they had ever seen each item of the test before. The full list of questions is available in Appendix B.5.

Accents Attitudes Scale (AAS). We created a 9-item survey to measure participants' explicit attitudes toward accents based on Contemori & Tortajada (2020) and Weatherholtz et al. (2014). The survey is available in Appendix B.6.

Implicit Association Test (IAT). The detailed description and the results of the test are available Appendix C.1 and will be available as a pre-print online together with three replications of the result later on.

After the surveys, the participants were fully debriefed. The experiment took about 30 minutes to complete.

4.2.4 Data analysis

All analyses were done in R (4.3.2) (R Core Team, 2020). R scripts together with the raw data are available on OSF. Individual difference predictors obtained in surveys (W-P, AAS, BFI) were centered and standardized (by subtracting the mean and dividing by the standard deviation). For all predictors, we fitted linear mixed-effects regression models using the *lmer()* function from the lme4 (1.1–35.1) (Bates et al., 2023) and *lmerTest* (3.1-3) (Kuznetsova et al., 2020) packages. Visual inspection of the residual plots showed no obvious violations of normality. Variance inflation factors were obtained using the *car* package (3.1-2) (Fox & Weisberg, 2018). The results were plotted with the *sjPlot* package (2.8.9) (Lüdecke et al., 2023).

The random structure was chosen by fitting a series of progressively complex models and comparing them using likelihood ratio tests with the *anova()* function. The base model included two crossed random intercepts for subject and item. By-participant random slopes were tested but produced a singular fit, indicating an overly complex model not supported by the data. Initially, all models contained three control variables from Cardillo et al. (2010) that significantly differed between the metaphor conditions: naturalness, figurativeness, and imageability. However, the variance inflation factor indicated high multicollinearity for figurativeness (VIF = 6.2), hence it was removed from potential predictors (note, however, that its presence or absence did not lead to any substantial changes in the results). This ensured low to modest collinearity for all the predictors (VIFs < 3.3).

The data was analyzed in a two-step procedure. In the first step, we fitted a model that contained the manipulated variables of Speaker (native, non-native), Metaphor Condition (familiar

metaphors, unfamiliar metaphors, literal), and their interaction. In the second step, we analyzed individual differences. Since some of them were moderately correlated and to prevent inflating the effects by fitting overly complex models, we fitted a separate model for each individual difference and its interaction with Speaker x Metaphor Condition (so, each model had ID x Speaker x Metaphor Condition as fixed effects, control variables, and the random structure). We then compared the baseline model with every new model using the *anova()* function. The detailed description of the calculation of individual difference scores, their distribution, and reliability is available below. The full correlation matrix is available in Fig. 4.2.

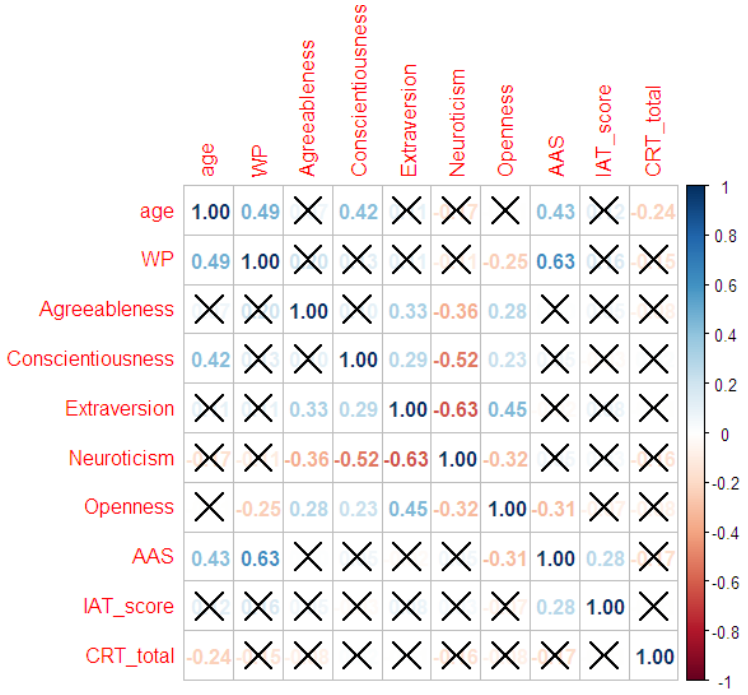

Figure 4.2. Correlation matrix for the individual difference measures. Black crosses indicate non-significant correlations.

Language Background. Mean age when the participants started learning English was 0.2 (range: 0–4, SD: 0.8). Fifteen participants (18%) were bilingual (defined as "native in two languages"). In our analysis, we only focused on bilingualism, the linguistic diversity of the participant's current and childhood environments, and the self-reported amount of interactions with non-native speakers, hence we will only discuss the correlations between those variables. The linguistic diversity of the participant's current environment was significantly correlated with that of their

childhood environment ($r = 0.40$, $p < 0.001$) but neither was significantly correlated with the self-reported amount of interactions with non-native speakers (both ps > 0.1).

Wilson-Patterson Conservatism (W-P) score. Higher scores indicate higher conservatism. Mean W-P score was 54/100 (range 30 to 85, SD = 16.5). Cronbach's α was excellent (0.90). The distribution of scores was not normal according to the Shapiro-Wilk's test (W = 0.92, $p < 0.001$). Visual inspection showed that the scores were aggregated around two clear peaks, 35-40 and 70-75. This is expected since, as mentioned above, the participants were recruited according to quotas.

Big-Five Personality Inventory (BFI). Cronbach's α was 0.90 for Extraversion, 0.69 for Agreeableness, 0.83 for Conscientiousness, 0.93 for Neuroticism, and 0.83 for Openness. Usually 0.7 is taken as an acceptable threshold (Cronbach, 1951), and values between 0.6 and 0.69 are considered questionable. Since we only analyzed Extraversion and Openness as predictors in this study, this was not an issue.

Accents Attitudes Scale (AAS). Higher scores indicate higher accent intolerance. Cronbach's α for the original questionnaire was 0.76, which is acceptable. Since it was a novel survey, we computed a correlation matrix to identify items that might not correlate with other items or with the total score using the 0.3 threshold. Although no items were correlated with the total score below that value, questions 3 and 4 were borderline (0.38 and 0.33). The principal component analysis suggested a presence of two separate components, PC1 containing questions 7, 2, 5, 8, 6 and PC2 questions 9, 3, 1, 4. Splitting the questionnaire into two components, however, only worsened the α for each component (0.75 and 0.63, respectively). Removing questions 3 and 4 barely improved α to 0.79 and did not change the results of the model where it was used as a predictor for sensibility ratings. Hence we decided to leave the survey as it is. Mean AAS score was 41/90 (range 15 to 75, SD = 13.3). The scores were normally distributed (W = 0.99, $p = 0.5$).

Cognitive Reflection Test (CRT). Eighty-seven percent of participants claimed to know 2 or more of the decoy questions (overclaiming exposure). This could potentially be explained by the fact that participants on Prolific regularly participate in experiments for money and may thus be familiar with different variations of this test. Notably, Stieger & Reips (2016) showed not only that experience with the questions may lead to higher scores, but also that experience with the CRT test is very confounded with education (particularly ceiling and floor scores). As a result, it is common practice to exclude all participants who claim to know the majority of the questions.

Removing all participants (N = 67) who claimed to know three or more questions of the survey, including the decoys, in this study would leave us with only 18 participants, rendering the data from this questionnaire unusable. We additionally ran a model predicting total CRT scores by familiarity, and it showed that the effect of familiarity was not significant (t = 0.53, p = 0.598). The highest degree obtained by the participant did not predict total CRT scores either (all ps > 0.3). We did find an effect of getting a Master's degree or higher on familiarity itself (t = 2.26, p = 0.026), in an expected direction (people with a Master's degree or higher scored on average 1.5 points higher on familiarity). The total scores were also significantly predicted by age (younger participants scored higher than older ones). We thus decided to proceed with analyzing CRT scores as a predictor without excluding participants for familiarity reasons. Instead, we put familiarity as a control variable.

For reaction times (how long it took the participant to read a sentence and make a rating), we first removed responses below 1000 ms and above 7000 ms based on the distribution and thickness of the tails (this eliminated 14.0% of responses). A square root transformation was applied to the resulting RTs (4683 data points), which yielded a near-normal distribution (skewness = 0.029). We then regressed the transformed RTs against the same predictors as in the main model for ratings. The effects of individual differences on RTs were not examined.

4.2.5 Results

Comprehension accuracy. Mean comprehension accuracy for the three questions following the written vignette was high: 94% across both blocks (91% in Block 1, 96% in Block 2). Comprehension differed significantly between blocks 1 and 2 (t = -2.55, p = 0.012) but did not differ between native and non-native blocks (t = 0.71, p = 0.5) according to paired t-tests. It is most likely that the participants, although they were told about the upcoming questions after the vignette, paid less attention to all the necessary information in the first block. Since the questions were the same (in a different order) after the vignette in Block 2, it is not surprising they did better.

Speaker evaluations. Post-experiment speaker evaluations did not differ between the two Speaker conditions according to paired t-tests. This was true for both overall evaluation ("How do you feel about this person overall?") (t = -0.11, p = 0.9, mean difference = -0.01) and likeability ("How likable do you think they are as a person?") (t = 1.55, p = 0.126, mean difference = 0.20). These

results are generally consistent with Bazzi et al. (2022) who found no differences in speaker ratings for affect, status, or solidarity using a written design. Of note, however, is the fact that we did find a strong prejudice against non-native speakers in the knowledgeability dimension using the Implicit Association test, when the participants could not consciously control their responses. The general cognitive category of foreign accents was associated with lower knowledgeability than the category of native accents according to the test results. Although the dimensions of evaluation were different, it may still suggest that the participants respond in a "socially acceptable way" in explicit surveys (or may not even be consciously aware of their biases, see Hewstone et al., 2002). We chose knowledgeability as it seemed the most relevant dimension for metaphor production and comprehension; however, Roessel et al. (2018) found a marginal difference for the dimension of affect in an IAT test as well.

Ratings and RTs. Mean sensibility ratings and response times for every experimental condition are reported in Table 4.2. The metaphor conditions worked as intended and showed the expected rating distributions, with literal sentences rated the highest and sentences with unfamiliar metaphors rated the lowest.

Table 4.2. Mean sensibility ratings and RTs for all experimental conditions (SDs in parentheses).

Metaphor Condition	**Speaker**			
	Rating		**RT (in ms)**	
	native	**non-native**	**native**	**non-native**
literal	44.0 (9.8)	43.8 (10.1)	3329 (1376)	3412 (1397)
familiar metaphors	37.9 (14.1)	36.6 (14.8)	3556 (1362)	3712 (1443)
unfamiliar metaphors	18.8 (17.2)	17.9 (16.8)	4036 (1293)	4081 (1303)

For linear-mixed models, both factorial predictors were sum-coded (Speaker and Metaphor Condition) so that metaphor familiar represented the difference between familiar metaphors and the grand mean, while metaphor unfamiliar represented the difference between unfamiliar metaphors and the grand mean. The results from the main model with ratings as a dependent variable and only the manipulated and control variables as predictors are reported in Table 4.3. Sensibility ratings differed significantly across Speaker and Metaphor Conditions. Interestingly, all sentences attributed to a non-native speaker were rated as less sensible compared to a native speaker. No two-way interactions were significant. In addition, naturalness affected ratings in a predicted direction: more natural sentences were rated as more sensible than less natural ones.

Table 4.3. Parameter estimates for a mixed-effects regression model predicting sensibility ratings. Model's formula: rating ~ Speaker*Metaphor Condition + naturalness + imageability + (1|participant) + (1|item). Conditional $R^2 = 0.548$. Significant terms are in bold.

Predictors	*β*	*CI*	*t*	*p*
(intercept)	**13.30**	**9.97 – 16.63**	**7.83**	**<0.001**
non-native speaker	**-0.40**	**-0.72 – -0.08**	**-2.48**	**0.013**
familiar metaphors	**1.24**	**0.46 – 2.02**	**3.13**	**0.002**
unfamiliar metaphors	**-5.18**	**-6.68 – -3.68**	**-6.78**	**<0.001**
naturalness	**4.58**	**3.92 – 5.23**	**13.65**	**<0.001**
imageability	-0.06	-0.66 – 0.53	-0.21	0.834
non-native speaker × familiar metaphors	-0.22	-0.70 – 0.25	-0.92	0.358
non-native speaker × unfamiliar metaphors	-0.03	-0.51 – 0.44	-0.13	0.898

We additionally examined the data for floor and ceiling effects. Strong ceiling effects were observed in the literal condition, with 41% of responses at the maximum value of the scale. This may have reduced the speaker effects by compressing the variability of responses. For familiar metaphors, 23% of responses were at ceiling, whereas 15% of unfamiliar metaphors were at floor.

The model with reaction times as a predictor is provided in Table 4.4. Overall, the results are consistent with the ratings, showing the same main effect of Speaker. It took the participants longer to read and evaluate sentences attributed to the non-native than the native speaker. Naturalness was significantly negatively correlated with RTs, so that more natural sentences predictably required less time to evaluate. Overall, there was a negative correlation between ratings and RTs ($r = -0.22$, $p < 0.001$), indicating that slower RTs were associated with lower ratings.

Table 4.4. Parameter estimates for a mixed-effects regression model predicting reaction times. Model's formula: sq.rt ~ Speaker*Metaphor Condition + naturalness + imageability + (1|participant) + (1|item). Conditional $R^2 = 0.433$. Significant terms are in bold.

Predictors	*β*	*CI*	*t*	*p*
(intercept)	**72.15**	**69.10 – 75.20**	**46.38**	**<0.001**
non-native speaker	**0.40**	**0.12 – 0.68**	**2.77**	**0.006**
familiar metaphors	**1.09**	**0.43 – 1.76**	**3.21**	**0.001**
unfamiliar metaphors	**-2.24**	**-3.52 – -0.95**	**-3.40**	**<0.001**
naturalness	**-2.78**	**-3.34 – -2.22**	**-9.70**	**<0.001**
imageability	0.08	-0.43 – 0.59	0.29	0.768
non-native speaker × familiar metaphors	0.26	-0.15 – 0.68	1.25	0.212
non-native speaker × unfamiliar metaphors	-0.16	-0.58 – 0.26	-0.77	0.442

We will now go over the individual differences.

Language and general background. When analyzing individual differences, we first considered each participant's language background. We started by adding the self-reported amount of interactions with non-native speakers as a predictor to the main model, but it did not yield a significant improvement in the model's fit ($\chi2 = 10.85$, $p = 0.093$). Next, we examined the participant's self-reported childhood language diversity (described as the number of people with foreign accents), but it also failed to improve the model ($\chi2 = 6.75$, $p = 0.344$). Including the self-reported language diversity of the participant's current environment marginally improved the model ($\chi2 = 11.93$, $p = 0.063$). The new model had a significant two-way interaction with the Speaker condition ($\beta = 0.43$, $CI = 0.12 – 0.73$, $t = 2.77$, $p = 0.006$). As shown in Fig 4.3A, while the ratings of sentences attributed to native speakers did not interact with the environment diversity, the ratings of sentences attributed to non-native speakers did. The higher the current environment diversity, the higher the participants rated sentences attributed to non-native English speakers[13].

Because of the prior literature showing the effect of bilingualism on metaphor processing, we additionally examined the effect of being bilingual. Bilingualism significantly improved the baseline model ($\chi2 = 24.0$, $p < 0.001$). Analyzing the model with bilingualism further, it turned out that it significantly interacted with both metaphor conditions (familiar metaphors: $\beta = 1.30$, $CI = 0.06 – 2.54$, $t = 2.05$, $p = 0.040$; unfamiliar metaphors: $\beta = -2.79$, $CI = -4.03 – -1.55$, $t = -4.41$, $p < 0.001$). As shown in Fig. 4.3B, bilingual participants rated both literal sentences and familiar metaphors higher than monolingual participants but, surprisingly, rated unfamiliar metaphors much lower than their monolingual peers. This contradicts previous research showing that multilingual speakers are more likely to derive figurative meaning from novel metaphors (Werkmann Horvat et al., 2022). Notably, only 18 percent of our participants self-identified as bilingual, so this effect is worth replicating with a larger participant sample.

[13] Importantly, although adding the self-reported amount of interactions with non-native speakers did not significantly improve the model's fit, the two-way interaction between Speaker and non-native speaker interactions was nevertheless significant as a term in the model ($t = 2.41$, $p = 0.016$) and, when plotted, turned out to have the exact same effect on ratings as the participant's current language diversity.

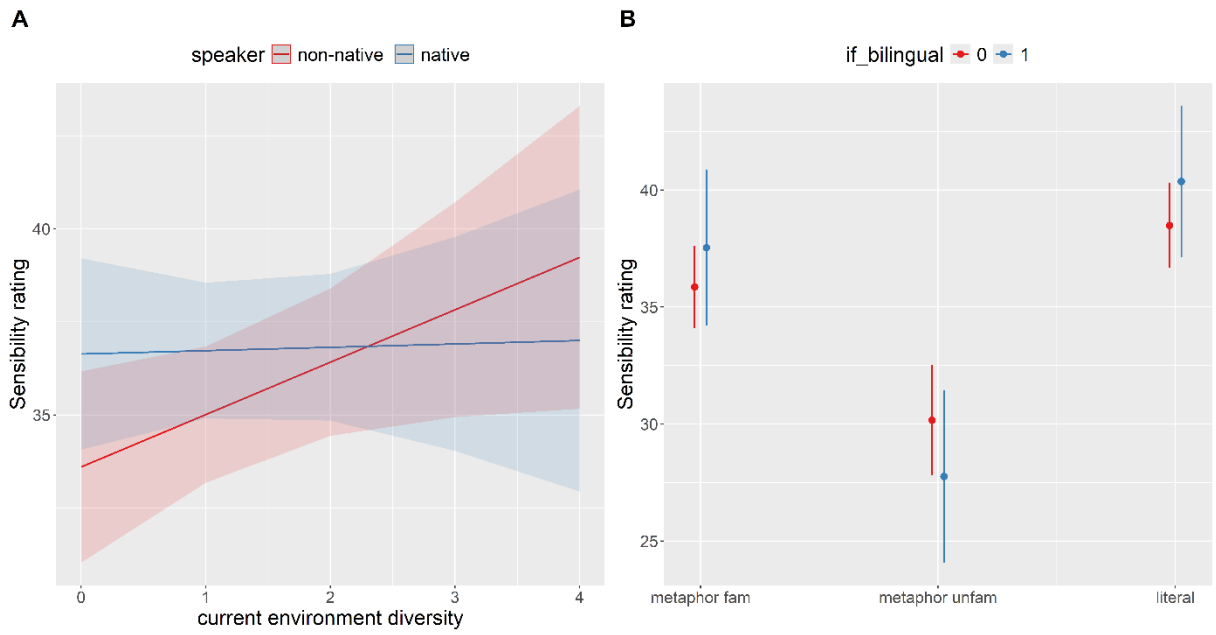

Figure 4.3. Predicted values (marginal effects) of the interaction between current environment diversity score and speaker (A) and between being bilingual and metaphor condition (B). The dependent variable is sensibility rating.

Cognitive Reflection Test. CRT scores significantly improved the base model ($\chi2 = 82.72$, $p < 0.001$). The new model revealed a (marginally) significant two-way interaction with Speaker ($\beta = 0.21$, $CI = 0.00 – 0.41$, $t = 2.01$, $p = 0.045$) and a significant two-way interaction with unfamiliar metaphors ($\beta = -1.14$, $CI = -1.44 – -0.84$, $t = -7.45$, $p < 0.001$). Both interactions are shown in Fig. 4.4. As can be seen in Fig. 4.4A, ratings for the non-native speaker consistently improved with higher CRT scores, whereas ratings for the native speaker remained unaffected. Fig. 4.4B shows that, whereas sensibility ratings for literal sentences and familiar metaphors were positively correlated with CRT scores, ratings for unfamiliar metaphors exhibited a negative correlation with CRT scores.

Wilson-Patterson Conservatism[14]. W-P scores significantly improved the baseline model ($\chi2 = 22.38$, $p = 0.001$). There was a main effect of W-P scores ($\beta = -1.44$, $CI = -2.65 – -0.24$, $t = -2.35$,

[14] Accents Attitudes Scale scores also significantly improved the base model ($\chi2 = 14.44$, $p = 0.025$), but the effect they exerted on sensibility ratings was indistinguishable from that of political leaning due to a high correlation between the two ($r = 0.63$, $p < 0.001$). Compared to the model with AAS scores, the model with W-P scores had a slightly lower AIC (42234 vs. 42242) and BIC (42346 vs. 42354). We thus do not report the results of the model with AAS scores in the main text (but the model and the plotted results are available in Appendix C.2).

$p = 0.021$) qualified by a two-way interaction with Speaker ($\beta = 0.39$, $CI = 0.07 – 0.71$, $t = 2.37$, $p = 0.018$) and a three-way interaction with Speaker and unfamiliar metaphors ($\beta = 0.60$, $CI = 0.12 – 1.08$, $t = 2.47$, $p = 0.014$). Fig. 4.5A shows that the native speaker was rated similarly regardless of Metaphor Condition–participants with lower Conservatism scores rated native speakers higher than participants with higher Conservatism scores. The non-native speaker was rated similarly except for the unfamiliar metaphor condition, where the participants' political orientation had no effect.

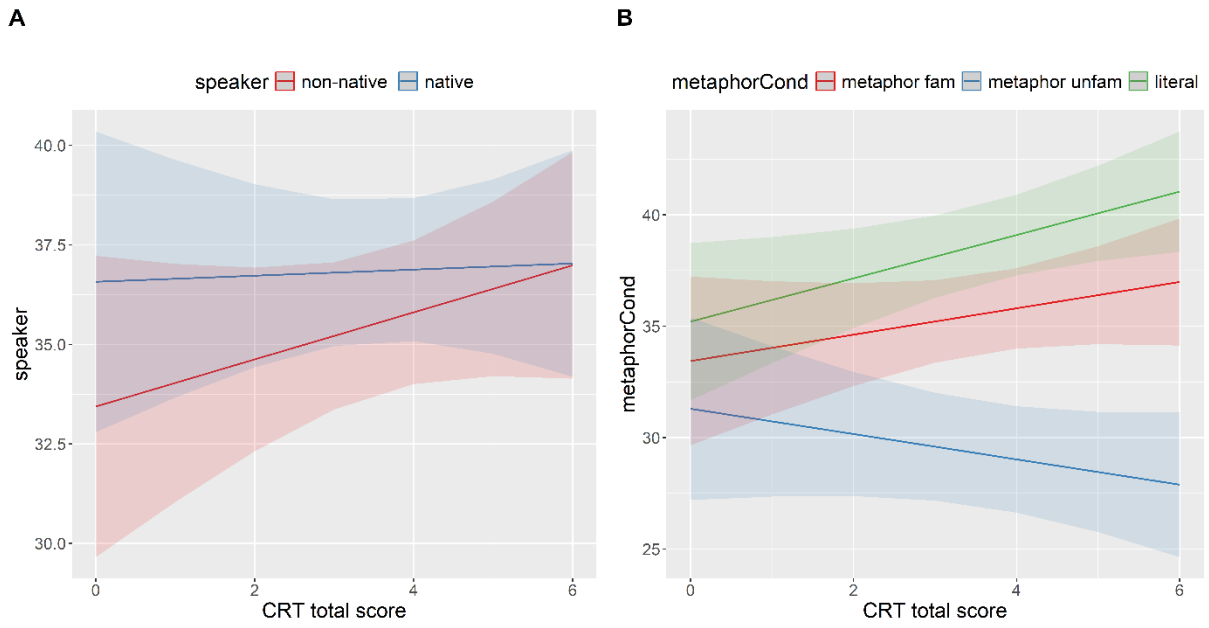

Figure 4.4. Predicted values (marginal effects) of the interaction between CRT scores and speaker condition (A) and between CRT scores and metaphor condition (B), with sensibility rating as a dependent variable.

Big-Five Inventory. Openness did not improve the model's fit ($\chi2 = 5.78$, $p = 0.449$), but Extraversion did ($\chi2 = 33.3$, $p < 0.001$). There was a two-way interaction with unfamiliar metaphors ($\beta = 0.93$, $CI = 0.45 – 1.40$, $t = 3.82$, $p < 0.001$) qualified by a three-way interaction with Speaker and unfamiliar metaphors ($\beta = 0.57$, $CI = 0.10 – 1.05$, $t = 2.35$, $p = 0.019$). As can be seen in Fig. 4.5B, while native unfamiliar metaphors were rated similarly regardless of Extraversion, non-native novel metaphors were rated significantly higher the more extraverted the participant was.

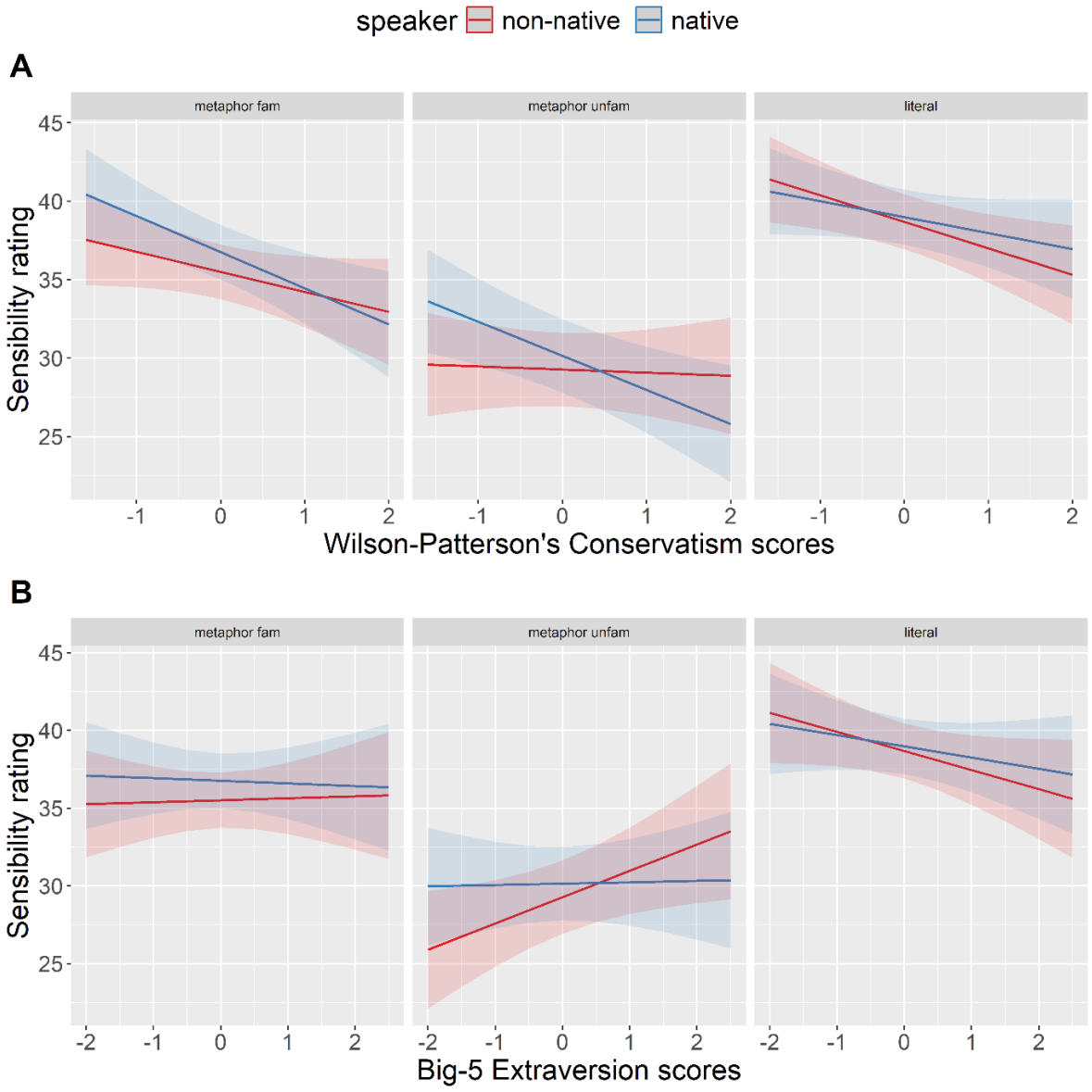

Figure 4.5. Predicted values (marginal effects) of the interaction between speaker condition, metaphor condition, and political orientation (A) and Big-5 Trait Extraversion scores (B), with sensibility rating as a dependent variable. Higher scores indicate higher conservatism, and higher extraversion.

IAT. Although IAT scores significantly improved the model's fit ($\chi 2 = 25.96$, $p < 0.001$), no significant interactions with Speaker were observed, hence we describe the model's results in Appendix C.2 only.

4.3 Discussion

The main goal of this study was to extend previous research on speaker-induced expectations in semantic and pragmatic processing by examining metaphorical sentences. Since both speaker (Bergen & Grodner, 2012; Bornkessel-Schlesewsky et al., 2013; Caffarra et al., 2018; Pélissier & Ferragne, 2022; Van Berkum, van den Brink, et al., 2008) and listener (Eekhof et al., 2021; Hammond-Thrasher & Järvikivi, 2023; Hubert Lyall, 2019; Mayn & Demberg, 2022; Puhacheuskaya & Järvikivi, 2022; Van Berkum et al., 2009b) effects in language processing are attested, we hypothesized that perception of metaphorical sentences may change depending on what is known about the speaker and on who the listener is. Specifically, because non-nativeness is prototypically associated with lower language proficiency, less agency over one's semantic choices, and a higher incidence of errors, we predicted that people would interpret metaphors differently if they believe that they come from L2 speakers.

We had two competing hypotheses for this study. According to *the leniency hypothesis*, metaphorical sentences from non-native speakers would be rated as more sensical since the readers would be more likely to derive a plausible interpretation from them and be more charitable in their judgments. According to *the downgrade hypothesis*, metaphorical sentences from non-native speakers would be rated as less sensical due to being perceived as unintentional awkward phrasing or semantic errors. The data lends support to *the downgrade hypothesis*. All sentences were consistently perceived as less sensical when attributed to an immigrant speaker with a strong non-native accent, mirroring the "illusion of ungrammaticality" reported for grammatical errors (Wesołek et al., 2023). Furthermore, these sentences also took significantly longer to process and evaluate than those attributed to the native speaker. This means that incorporating the identity of the non-native speaker was also cognitively taxing. This contributes to mounting literature showing that the speaker identity plays a crucial role in sentence interpretation, sometimes in the earliest stages of processing (Grey & van Hell, 2017; Hanulíková et al., 2012; Van Berkum, Van Den Brink, et al., 2008). Crucially, this judgment occurred in the absence of any oral speech—the participants were merely informed about the speaker's accent through a written vignette. Also importantly, the vignette did not mention anything about the speaker's language proficiency. This indicates that the lower ratings were driven by preconceived notions about non-native speakers'

linguistic competence rather than the extra cognitive load associated with foreign-accented speech processing.

Importantly, we did not find any significant interactions with the Speaker condition. This aligns with the findings of Gibbs et al. (1991) where participants rated metaphorical and literal sentences as less meaningful when they assumed that they were created by a computer program rather than a human poet. The lack of difference between the figurative and the literal condition also agrees with the findings of Bazzi, Brouwer, & Foucart (2022). Using a strictly written modality like the present study, Bazzi and colleagues found that both literal and ironic sentences attributed to non-native speakers were perceived as less ironic than those attributed to native speakers. Thus, expectations induced by merely describing someone as foreign-accented may colour subsequent interpretation of that person's language regardless of the particular linguistic phenomenon.

It is also important to address the magnitude of these effects. Since metaphorical language has not been investigated in this paradigm before, direct comparisons with previous research are not possible. However, the difference in ratings between the two speaker conditions (1.4 for familiar metaphors and 1.1 for unfamiliar metaphors on a 50-point scale; 0.40 overall including literal sentences) is proportionally of a similar magnitude with the effects reported for written designs in previous studies (0.11 on a 7-point scale for irony in Bazzi, Brouwer, & Foucart, 2022; 0.04 on a 5-point scale in Experiments 1B and 2B and 1.5 on a 100-point scale in Experiment 3 in Lorenzoni et al., 2022, 0.09 on a 5-point scale in Fairchild & Papafragou, 2018). Overall, while these effects may be modest, the methodology appears to consistently yield results. This is important for the field since written presentation allows researchers to completely eliminate any effects of cognitive load due to speech disfluency and thus better isolate the effects triggered by foreignness and lower L2 command.

That said, a more detailed discussion of the issue of modality and associated cognitive load is merited. The written paradigm for exploring the effects of non-native accents pioneered by Fairchild and Papafragou (2018) is becoming increasingly popular as it allegedly allows for an examination of non-nativeness in isolation from processing confounds. Whether that is really the case is, in fact, unclear. To begin with, research has shown that readers generate implicit prosody (inner speech) when reading, and that this implicit prosody affects the processing of written

sentences (see Breen, 2014 for a review). In addition, studies on auditory perceptual simulation have shown that when the readers are directly asked to imagine the native or non-native speaker's voice while reading the sentences, their reading speed is affected by the speech rate of the imagined voice (but not the identity of the speaker per se) (Zhou & Christianson, 2016b, 2016a). Importantly, in a follow-up study, Zhou et al. (2019) showed that readers exhibit attenuated N400 and P600 components when they silently simulate a non-native speaker voice when reading ungrammatical sentences. Although the authors argued that such perceptual simulations lead to a more detailed prosodic representation of the text than the default prosodic contour, this has yet to be empirically tested. It is theoretically possible that readers might unconsciously generate "non-native" implicit prosody even without the cued simulation, especially when some accent is played before each block or even each trial (as in Zhou et al. 2019). This perceived non-fluency could affect the way readers process the sentence cognitively. More research on the topic is warranted.

Another important consideration is whether the results are driven directly by the lower expected language proficiency or indirectly by the accent prestige. It is a known fact that accents carry varying levels of prestige (Giles, 1970, 1973). If an accent carries lower prestige, readers may make additional assumptions about the speaker, such as lower level of education, which would indirectly contribute to the effect found in this study. That said, both native and non-native speakers were introduced as undergraduate students from the same university, which should have reduced the influence of such stereotypes, if any. We chose a Chinese accent because of its high prevalence in North America (hence higher familiarity of the participants with that accent). The Canadian accent, as discussed in the Methods section, was chosen to make the native speaker an immigrant together with the non-native speaker. We were not able to find any studies exploring how prestigious Canadian accents are considered in the U.S. and whether it varies by state. If our results were indeed influenced by accent prestige, outcomes might differ when a different native or non-native accent is used. This is an important empirical question which should be addressed in follow-up research.

Tangentially related to the above discussion are the results of the Implicit Association Test. This is our fourth replication of this result, now extended to the American population. The huge size of the effect suggests that the subconscious bias against speakers with foreign accents is very strong and shows that lower competence of non-native speakers is generally expected. It is unclear

whether this implicit bias may reinforce other assumptions about the non-native speaker, particularly if some of them are driven by accent prestige, and whether those assumptions, in turn, contribute to the overall result. The IAT test scores did not directly interact with Speaker condition in this experiment. One possibility is that actual foreign-accented speech might be needed to see any effect of IAT scores, and merely presenting the non-native speaker through a written vignette may not be sufficiently effective. Since this measure has almost never been used as a predictor in psycholinguistic studies, more research is needed to make any conclusions. It is important to note, however, that explicit evaluations of the speakers *themselves* ("How do you feel about this person overall?" and "How likable do you think they are as a person?") did not differ between the native and non-native speakers. This lack of difference in explicit evaluations could be influenced by social desirability bias, where participants are reluctant to express negative attitudes toward non-native speakers.

Perhaps more importantly, not everyone had the same response to the experimental conditions. Several individual differences had a significant effect on ratings, either independently or in interaction with other conditions. Of primary interest to us were variables that tapped into non-native accent exposure, explicit and implicit language attitudes, sociality, and intelligence (in particular, cognitive reflection, i.e. the ability to override an incorrect intuitive response). We found that some of those variables were indeed predictive of participants' ratings. To briefly summarize our findings, the tendency to judge all or some non-native sentences as less sensical than those by native speakers mostly came from people from less linguistically diverse backgrounds, those who were less extraverted, and those who had lower cognitive reflection. We will discuss these findings in more detail below.

One notable finding was the effect of environmental diversity. The participants who self-reported lower levels of linguistic diversity in their current work or school setting rated all stimuli attributed to non-native speakers lower than those who reported a more diverse environment, whereas ratings for native speakers remained unaffected. This contributes to previous research showing that individuals with greater exposure to different accents have more positive attitudes toward non-native speakers and that such positive attitudes predict better comprehension accuracy (Dewaele & McCloskey, 2015; Ingvalson et al., 2017). Greater accent exposure has also been linked to less disruption of prediction when listening to foreign-accented speech (Porretta et al., 2020). However,

it is also important to acknowledge the potential for a "chicken or egg" dilemma here. As argued by Dragojevic et al. (2017), the disruption in cognitive operations by disfluency may lead to negative attitudes independent of the content being processed, either directly or indirectly via affect. It is thus possible that more difficulties with processing a particular accent during one's lifetime may foster more negative affect/attitudes toward that accent, which in turn could exacerbate difficulties in processing it, creating a self-reinforcing loop between negative attitudes and comprehension challenges. Nevertheless, the findings in the present study seem intuitive and in line with our expectations. It is likely that higher "leniency" that people with high diversity scores demonstrated to non-native speakers is mediated by more positive attitudes toward non-native speakers in general. Methodologically, it is important to note that this self-reported amount of linguistic diversity was a stronger predictor than the self-reported amount of interactions with non-native speakers or the diversity of the participant's childhood environment. The reasons for this are not entirely clear. Although we put in a lot of effort to get a representative sample, the scores for all these measures were somewhat clustered in the lower end of the scale, particularly the scores for the diversity of the participant's childhood environment, which likely influenced which predictors came out stronger. Replicating this effect with a more representative sample, perhaps using quotas to ensure a wider range of diversity experiences, would be beneficial.

Another individual variable that predicted ratings was cognitive reflection. Cognitive reflection scores had a marginally significant effect on ratings for each speaker and a significant effect on ratings for unfamiliar metaphors. Prior research has linked greater cognitive reflection scores with a better ability to draw pragmatic inferences (Mayn & Demberg, 2022)[15]. In our study, participants with higher cognitive reflection scores gave higher sensibility ratings for literal sentences and familiar metaphors but rated unfamiliar metaphors as less sensible. Thus, in a nutshell, higher cognitive reflection scores lead to more critical evaluation of unconventional language and better evaluation of conventional language. One might ponder why this would be the case. Generally speaking, individuals with higher cognitive reflection scores tend to be more skilled at overriding immediate, surface-level interpretations and engage in further thinking. It would thus seem that

[15] It is noteworthy that CRT scores were highly correlated with IQ measures in Mayn & Demberg (2022), hence it was unclear whether or not the effects of the two could be separated. We did not collect IQ scores; hence we cannot make any conclusions about whether CRT scores in our study contributed beyond other measures of intelligence. Of note, CRT scores in our study were not predicted by the participant's highest level of education.

scrutinizing novel metaphors further resulted in lower acceptance of them. Additionally, lower cognitive reflection was marginally positively associated with evaluation of sentences attributed to non-native speakers—but not native speakers. Future research should explore this finding further, perhaps with actual foreign-accented speech as well.

Extraversion also predicted ratings in a unique way, so that higher extraversion was associated with higher ratings of non-native unfamiliar metaphors. To probe this finding further, we checked the correlation strength between Extraversion and frequency of interactions with native and non-native speakers. To our surprise, while Extraversion was significantly positively correlated with the self-reported amount of non-native speaker interactions ($r = 0.25$, $t = 19.1$, $p < 0.001$), it was actually negatively correlated with the amount of native speaker interactions ($r = -0.14$, $t = -10.1$, $p < 0.001$). Although the latter correlation likely reflects a tendency for extraverted participants to underestimate their interactions with native speakers, this finding is still noteworthy. Although we did not predict this finding in particular, we expected Extraversion to be correlated with higher ratings for metaphorical sentences and possibly with higher ratings for metaphorical sentences attributed to non-native speakers. Thus, this result is in line with our general expectations. One possible explanation for this finding may be that extraverts are generally more socially active and engage more with others, including people from different cultural and linguistic backgrounds, which may make them less bothered by accents. In addition, more social interaction may lead to increased exposure to different linguistic constructions, which again may make extraverted people more lenient in their judgments. As Extraversion was positively associated with the amount of interactions with non-native speakers in our sample, this supports our explanation.

Finally, we found an effect of political orientation. Although not often researched in psycholinguistics, political ideology has been previously found to predict processing of different linguistic stimuli, in particular those that are socially charged (Hammond-Thrasher & Järvikivi, 2023; Hubert Lyall, 2019; Puhacheuskaya & Järvikivi, 2022). Unlike prior studies that relied on university undergraduates, we recruited a representative sample through Prolific, with a preset quota of 50 percent conservative and 50 percent liberal participants. We obtained a very good distribution of scores in the Wilson-Patterson Inventory that we administered to quantify political orientation, with two clear peaks in the first and the second halves of the scale. Importantly, the scores significantly interacted with the experimental conditions. Overall, less conservative

participants rated almost everything significantly higher than their more conservative peers. It would thus appear that more liberal political leaning is associated with a more lenient judgment style and "positive evaluation bias". There have been previous reports of this bias in more left-leaning participants (Hubert Lyall, 2019). This bias could not be explained by, for instance, greater Openness that was found to be correlated with political orientation (Sibley et al., 2012) because we specifically examined that Big-5 trait and it was not predictive of ratings by itself or in any experimental condition. The only exception to this positivity bias were non-native unfamiliar metaphors that were rated the same regardless of the participant's political orientation. Why only non-native unfamiliar metaphors were unaffected by political leaning is not entirely clear. One speculative explanation is that, since these metaphors are so unconventional, participants might approach them with a similar level of uncertainty, regardless of their political beliefs. However, it is unclear in this case why native unfamiliar metaphors did not exhibit the same trend. Since we did not predict this finding, caution is advised when evaluating our explanation, and more research is warranted. Additionally, it is not entirely clear why the three-way interactions in this study are limited to unfamiliar metaphors. One possibility is that unfamiliar metaphors exhibited the largest rating variability (as indicated by the highest standard deviation) hence individual differences come into play stronger.

Follow-up research could use more time-sensitive methodology like EEG or eye-tracking to provide more nuanced insights into the effects found in this study. It would be useful to know whether the interpretation difficulty arises already at the verbal metaphor, with those attributed to non-native speakers appearing as more anomalous (as evidenced by the N400 component or longer processing times). Alternatively, it may be that the effect only arises at the sentence level (i.e. in ratings or other decision-making tasks). Our design makes it impossible to disentangle different explanations. In addition, follow-up research could explicitly indicate the L2 proficiency of non-native speakers before the experiment and examine whether the effects found in this study differ with low versus high explicit L2 proficiency.

The results of our study do not support or reject any particular psycholinguistic model of metaphor processing. As for the models of language processing in general, there have been attempts to adjust them to allow for constraining influence of factors associated with the speaker, listener, or situation where communication takes place (Kidd et al., 2018; Münster & Knoeferle, 2018; van Berkum,

2018), since mounting evidence shows the influence of such factors even on early stages of language comprehension. Our data aligns with those adjusted models.

Overall, our findings demonstrate that preconceived notions about non-native speakers significantly affect how sensible their sentences are perceived. Additionally, they contribute to the literature showing the effects of non-nativeness even when no oral speech is experienced. They also highlight interpersonal variation in ratings and the importance of taking that variation into account.

4.4. Ethics

All participants gave consent to participating in the experiment after they were provided with a full description of the study. The plan for this study was reviewed for its adherence to ethical guidelines by a Research Ethics Board at the University of Alberta (reference number Pro00134190).

4.5 Acknowledgements

This research was supported by a Social Sciences and Humanities Research Council of Canada Partnership Grant (Words in the World, 895-2016-1008).

Chapter 5

Accents of trust: The impact of knowledge domain, group membership, and implicit accent attitudes on non-native speaker credibility

Abstract

Whether non-native speakers are considered less credible overall is a matter of debate, with mixed results across studies. Competing theoretical frameworks attribute potential credibility gap to either processing difficulty (fluency-based accounts) or social categorization (outgroup-based accounts). Although general bias against non-native speakers so far lacks empirical evidence, domain-specific credibility remains unexplored. In two experiments using different modalities, we examined whether non-native speakers are trusted less on facts related to "ingroup knowledge". To further probe the malleability of social categorization effects, we manipulated the group membership of foreign-accented speakers, introducing them as either Canadian citizens or Immigrants. Experiment 1 with written stimuli showed weak evidence that non-native speakers were indeed believed less when making statements of unknown truth value, regardless of their group membership. In contrast, in Experiment 2, participants were more inclined to believe *all* statements spoken by a non-native individual introduced as a Canadian citizen. This suggests a modality-specific "knowledgeable immigrant" effect that challenges intelligibility-based accounts. This effect was only observed in monolingual participants and absent for bilinguals. We also explored the role of listeners' implicit attitudes. Using the Implicit Association Test (IAT) with knowledgeability attributes, we found a significant bias against foreign-accented speakers in both studies. Crucially, IAT scores significantly modulated truth ratings for oral stimuli. The strong implicit bias against foreign-accented speakers, coupled with the complex interaction between accent and explicit group membership, suggests that the mechanism underlying accent-based veracity judgments is more sophisticated than a simple processing fluency account would predict.

Keywords: credibility, truth evaluation, accent, social categorization, individual differences

5.1 Introduction

One of the major social markers affecting communication is the speaker's linguistic identity manifested through their standard or non-standard (e.g., regional or foreign) accent. Research consistently demonstrates that standard-accented speakers are evaluated more favorably on most dimensions than those with non-standard accents, with effect sizes at the upper bounds in psychological research (Fuertes et al., 2012). Even in multicultural societies with high immigration rates, such as the United States, a thorough examination of print media shows that a foreign accent is consistently portrayed as an impediment to communication (Gluszek & Hansen, 2016; Lippi-Green, 2012). It is also often emphasized that foreign-accented speakers should invest significant effort to get rid of their accent, despite scientific evidence convincingly showing it to be near-impossible when the speaker is an adult (e.g., Birdsong, 2006; Dollmann et al., 2020; Piske et al., 2001; Scovel, 2000; Tahta et al., 1981, among others).

Since communication is a multifaceted process that goes far beyond the mere exchange of factual information, expectations induced by the speaker identity have been shown to profoundly affect how the message is perceived and interpreted (Bergen & Grodner, 2012; Pélissier & Ferragne, 2022; Regel et al., 2010; Van Berkum, van den Brink, et al., 2008). It is, therefore, unsurprising that a foreign accent has been found to influence linguistic processing on multiple levels. These effects, however, are not uniform. Perhaps the most significant contribution to processing comes from the acoustic signal itself. Substantial deviations from native phonology and prosody exhibited by L2 speakers (Gut, 2012; Hanulíková & Weber, 2012; Wester et al., 2007) are often cognitively taxing to process, which leads to more negative social evaluations (Dragojevic et al., 2017; Dragojevic & Giles, 2016). The situation is further complicated by the fact that certain phonetic features of non-native accents, such as the substitution of interdental fricatives with plosives (Wester et al., 2007), often overlap with features found in lower class dialects of English (Ryan, 1983).

Besides these acoustic challenges, non-native accents also induce expectations of lower language proficiency, including diminished control over semantic choices and syntactic instability. Furthermore, since it conveys crucial information about the speaker's background, accent is one of the most reliable markers for social categorization in diverse multicultural societies.

Monolingual listeners who speak standard language varieties may instantly categorize non-native speakers as an outgroup—not only different from themselves but also of lower societal status (Ryan, 1983). Some authors have argued that the outgroup status may even be a more influential factor than fluency in shaping perceptions (Gluszek & Dovidio, 2010; Mai & Hoffmann, 2014). Research has established that the human brain encodes race, gender, and social status with remarkable speed and automaticity (Ito & Bartholow, 2009; Ito & Urland, 2003; Rule et al., 2012; Zink et al., 2008), and, once this categorization occurs, people start treating their ingroup and outgroup differently. There is a general tendency to assume that ingroup members share your values, beliefs, and attitudes, and that outgroup members have a contrasting perspective (Dovidio & Gaertner, 2010). Consequently, social categorization frequently results in ingroup favoritism and/or outgroup antagonism (Ito & Bartholow, 2009; Molenberghs & Louis, 2018). The preference for native accents is particularly robust. It has been shown that adults consistently choose accent over visual ethnic cues for social categorization (Rakić et al., 2011). This bias is not limited to adulthood but emerges early in human development. Infants preferentially look at native speakers before they can even comprehend speech, while children choose native over non-native speakers as their friends (Imuta & Spence, 2020; Kinzler et al., 2007, 2009). In fact, accent preferences precede racial preferences in the course of a child's development (Kinzler et al., 2009; Kinzler & Spelke, 2011). This finding has led some authors to argue that accents have played a more important role than race during evolution, as neighboring communities were more likely to differ in accents rather than races (K. A. Collins & Clément, 2012).

Naturally, the impact of these accent-based expectations extends beyond social preferences. When it comes to language processing, numerous psycholinguistic studies with different methodologies have found significant differences between native and non-native speech. Semantic anomalies and common syntactic errors in non-native speech do not evoke the P600 component that is commonly taken to index reanalysis and repair (Hanulíková et al., 2012; Romero-Rivas et al., 2015) as compared to native speech. Pragmatic inferences can also be negatively affected (Caffarra et al., 2018; Puhacheuskaya & Järvikivi, 2022), although surprisingly, they can also be affected in non-native speakers' favor (Fairchild et al., 2020; Fairchild & Papafragou, 2018). Prediction has been shown to be impeded by the foreign accent, although the strength of this impediment decreases as a factor of accent familiarity (Porretta et al., 2016, 2020).

Perhaps one of the most surprising findings on the effects of non-native accents concerns credibility. Lev-Ari & Keysar (2010) were the first ones to report that statements of a generally unknown truth-value like "Ants don't sleep" are less likely to be rated as true when spoken by a non-native speaker. They also argued that prejudice against foreigners does not play a role in the "incredibility effect" because the non-native speakers in the experiment were said to be only reciting facts that a native experimenter gave them rather than expressing their own knowledge (i.e. fluency-based account). However, several issues with the study remain unresolved. First, these findings have been challenged by largely unsuccessful replication attempts, even when the authors used virtually identical methodology (Foucart et al., 2020; Foucart & Hartsuiker, 2021; Meo et al., 2011; Podlipský et al., 2016; Souza & Markman, 2013; Stocker, 2017; Wetzel, 2021). One recent study even found the opposite pattern (Lorenzoni et al., 2022). A study that used regional accents likewise found no effect of accent (Frances et al., 2018; cf. Pycha & Zellou, 2024). Although the original finding was later replicated on a bigger sample by the same author (Boduch-Grabka & Lev-Ari, 2021), the new study failed to reproduce in a close pre-registered replication attempt with the same audio materials by different researchers (Barlow et al., 2024). Second, when studies did find differences between native and non-native speakers, it was often in neurocognitive measures (ERP) and not behavioral ratings (Foucart et al., 2019; Foucart & Hartsuiker, 2021). Third, Foucart et al. (2019) did not use any *speech* at all—sentences were presented in the written form with merely a picture of the speaker (who was earlier introduced as either having or not having a foreign accent). Thus, processing difficulty was kept constant between the speakers, which means that any effects they found stemmed from the social categorization into ingroup/outgroup instead (i.e. outgroup-based account). The intelligibility-based framework becomes even less convincing in light of the fact that white noise, although it affects processing ease, does not make statements any less credible (Souza, 2012, p. 42). Furthermore, while Boduch-Grabka & Lev-Ari (2021) reported that previous exposure to foreign-accented speech improved subsequent veracity judgments of foreign-accented statements of unknown truth value, this pattern of results did not replicate (Barlow et al., 2024). Lastly, if we indeed trust non-native speakers less, it is not clear if this is because we consider them less knowledgeable (Roessel et al., 2018) or because we think that they are outright lying (Evans & Michael, 2014).

The studies reviewed above suggest that generalized prejudice against foreign-accented speakers as reflected in truth judgements may not stand to scrutiny. However, domain-specific prejudice has not been investigated before. Most prior studies combined very different kinds of world knowledge, such as political facts, biology, history, geography and politics of various countries, and inventions, as well as mixed in narrow, "expert" knowledge with common knowledge (e.g., Foucart & Hartsuiker, 2021; Hanzlíková & Skarnitzl, 2017; Lev-Ari & Keysar, 2010). Considering how many stimuli the researchers need to come up with for psycholinguistic studies, this is understandable. However, it can also be argued that it introduces confounds, since the perception of the speaker's knowledge has been shown to directly influence pragmatic inferences made by listeners (Bergen & Grodner, 2012). It is reasonable to assume that the credibility of foreign speakers may additionally vary depending on the topic. Immigrants, particularly recent ones, may be perceived as less credible than native speakers when it comes to "ingroup knowledge"—facts related to the native listener's country, such as geographical, political, or cultural knowledge. On the flip side, they may be considered more credible when it comes to "outgroup knowledge"—facts about their own country or other countries they have lived in (see Fairchild & Papafragou, 2018 for a similar idea). Most importantly, the credibility associated with the ingroup knowledge may evolve over time as immigrants become more integrated. For example, by the time somebody acquires new citizenship, they may potentially be perceived as equally knowledgeable with native speakers. For these reasons, our critical stimuli focused specifically on Canada (native country of our participants).

In addition, findings by Foucart and colleagues suggest that if social categorization makes a major contribution to the "incredibility effect", then it could be manipulated. Prior research has shown that some outgroup effects can be offset by activating (or creating) a different identity of the speaker that is shared with the listener (the so-called common in-group identity, see Gaertner et al., 1993). These groups can be meaningful or completely arbitrary (minimal), such as "dot overestimators" or "dot underestimators" following the counting of dots on the screen (Tajfel et al., 1971; Tajfel & Turner, 1986). For instance, Van Bavel et al. (2008, 2011) showed that after assigning people of different races to experimentally created groups, they show more fusiform face area activity for ingroup compared to outgroup faces regardless of their race. The authors also demonstrated that the observed effects are due to the increased neural activity for the ingroup faces

rather than decreased activity for the outgroup faces. This increased ingroup favoritism (rather than outgroup antagonism) is a converging finding across many studies (see Ito & Bartholow, 2009 for an overview of EEG studies and Molenberghs & Louis, 2018 for an overview of fMRI studies from different disciplines). Whether accent biases can be reduced by activating different shared identities remains an open question. Recent research on accent perception showed that presenting a foreign-accented speaker with an Immigrant/New Canadian label significantly increased accent strength rating in native Canadian listeners (Calkins et al., 2023). Importantly, presenting the same speakers with the Canadian label as compared to no label at all did not affect accent perception in that study. Ryan & Sebastian (1980) presented foreign-accented speakers as either being from low or middle-class backgrounds and found that middle-class accented speakers were rated higher than lower-class ones on almost every dimension except for speech intelligibility. Surprisingly, lower-class accented speakers were perceived as easier to understand than middle-class accented speakers. In line with these findings and crucial for the present study, Foucart et al. (2019) manipulated the social status of native and non-native speakers through a combination of auditory introductions and a "hierarchy game" and found that both accent and status significantly modulated world knowledge integration as indicated by the N400 and the P600 components (but not truth ratings). The authors reported a more negative N400 and a long-lasting negativity in the P600 window for unknown statements compared to true statements only for the high-status native-accented speaker (true and unknown sentences for the high-status foreign-accented speaker did not differ). The authors explained it by shallower processing of statements attributed to the foreign-accented speaker, possibly indicating that they were considered less reliable. All in all, explicit cues to social group membership seem to alter how non-native speakers and their speech are perceived.

5.1.1 The present study

The goal of the present study was threefold. First, we wanted to investigate how much of the "incredibility effect" could be attributed to the speaker's outgroup status versus the cognitive difficulty of processing foreign-accented speech. To tease apart intelligibility-based accounts from outgroup-based accounts, we compared written (Experiment 1) to aural (Experiment 2) delivery of the same stimuli. Second, considering that current evidence points toward the outgroup explanation of the credibility gap, if it exists (Foucart et al., 2019; Foucart & Hartsuiker, 2021),

we wanted to explore how malleable the effects of such categorization are. For that, we labelled the foreign-accented speakers as either Canadian or Immigrant during the introduction phase, thus making them either a partial ingroup/outgroup or a double outgroup. To improve authenticity, foreign-accented speakers who were introduced as Canadians said that they have lived in Canada for quite some time, whereas those introduced as Immigrants said that they have recently immigrated to Canada. Of course, some confounds between explicit labelling and length of residence are inevitable, since acquiring Canadian citizenship instantaneously without living here for a required period of time is generally impossible. That said, an average person has only a vague understanding of how long one needs to reside in Canada to obtain their citizenship, and the terms describing length of residence were very non-specific. Group membership was further solidified by showing a corresponding country flag, which remained visible under the speaker's photo throughout the trials (the label itself was no longer shown). Lastly, we conducted an exploratory investigation of whether the listener's implicit attitudes toward foreigners mediate world knowledge integration.

There are several reasons for choosing implicit over explicit attitudes. Most studies in psycholinguistics that investigate prejudice focus exclusively on explicit measures, like surveys, which are prone to self-presentation strategies due to ever increasing social condemnation of prejudiced views (Hewstone et al., 2002). This problem may be exacerbated during in-person experiments where being observed, or simply just being aware that the experimenter is present outside the experiment booth, may motivate participants to modify their behavior. They may give socially desirable answers out of fear of being judged or as part of the self-presentation strategy, even when explicitly told that their data will be anonymous and that they will not be identifiable in any reports. This phenomenon is known as the social desirability bias (Edwards, 1957; Krumpal, 2013). Sometimes people can be sincerely unaware that they hold such biases and how they affected their behavior at a particular time (Nisbett & Wilson, 1977). Some authors have, in fact, argued that social prejudices are nowadays just more covert and unconscious for an individual, which makes them more difficult to detect and control (Amodio, 2014). One way to detect covert biases is the Implicit Association Test (IAT) (Greenwald et al., 1998). The IAT does not involve explicit questionnaires and instead assesses the ease of association between a bipolar target concept (e.g., gay/straight) and a bipolar attribute (e.g., positive/negative). The core idea is that quicker

responses are expected when two highly associated concepts are mapped onto the same button (straight/positive and gay/negative) compared to different buttons (straight/negative and gay/positive). We decided to use the IAT as a predictor for truth evaluations in this study. To our knowledge, generalized implicit prejudice against foreign-accented speakers has never been used as a predictor in psycholinguistic studies before. In fact, we are aware of only one psycholinguistic study that used the IAT at all, which was done to test whether implicit Australia and New Zealand biases affected phonetic accommodation (Babel, 2010). The full design and procedure of the Implicit Association test is outlined in the Materials section.

Our hypotheses for the study were as follows. We predicted that unknown statements by non-native speakers will be judged as less truthful than those by native speakers. We did not expect any effects of accent on facts known to be true or false, and used them for control purposes only. In addition, considering that Foucart et al. (2019) found neurocognitive differences in response to native and non-native speaker statements without any actual speech and that Souza (2012) did not find any effect of fluency-disrupting noise on speaker credibility, we expected the effect to be present with both written and auditory delivery of stimuli, although we expected it to be weaker in the former case. Regarding the group membership manipulation, we hypothesized that unknown statements by a foreign-accented speaker introduced as Canadian would be rated as more truthful than those by a foreign-accented speaker introduced as Immigrant but still lower than those by a native speaker of Canadian English. This might be due to raising the social status of the speaker, due to indicating longer residency in the country, or both. Yet again, this effect might be stronger when oral stimuli are used. As to the individual differences in implicit prejudice, we theorized that participants with stronger prejudice against foreign-accented speakers (provided that this prejudice exists in the Canadian context, as it was never examined before) will exhibit harsher credibility judgments toward them and may even be less affected by the group membership manipulation.

Experiment 1

5.2 Materials and methods

5.2.1 Participants

Thirty-five self-identified native English speakers with no history of neurological or psychiatric conditions participated in the experiment. All participants were students and received 20 CAD per hour (N = 20) or partial course credit (N = 15) for their participation. All self-identified as Canadian. Five participants stated that they had come to Alberta to study within the past couple of years and had not lived here most of their life. Since a big proportion of our statements were Alberta-centric (one third of all the critical stimuli; see Stimuli), their data was not analyzed. The final sample thus consisted of 30 participants. Mean age was 20.5 (range: 18–28, SD: 2.3). Nineteen participants (63%) self-identified as female, 8[27%] as male, 3[10%] as non-binary.

5.2.2 Materials

5.2.2.1 Stimuli

One hundred and thirty-one sentence triplets were initially created. Each triplet had the following conditions: True (sentences that contain information known to be true), False (sentences that contain information known to be false) and Unknown (sentences that are factually correct, but this is unknown to the participants). We did not mix true and false facts into the unknown condition to maintain better control over the stimuli. Three example triplets are provided in Table 5.1. The full list of stimuli with their characteristics is available on OSF.

Table 5.1. Examples of experimental triplets. The target word is italicized.

Condition	Sentence
True	The most populated city in Canada is *Toronto*.
False	The most populated city in Canada is *Hamilton*.
Unknown	The tenth most populated city in Canada is *Hamilton*.
True	The last Olympics in Canada took place in *Vancouver*.
False	The last Olympics in Canada took place in *Toronto*.
Unknown	The last execution in Canada took place in *Toronto*.
True	The flag of Canada features a *leaf*.
False	The flag of Canada features a *star*.
Unknown	The flag of Nunavut features a *star*.

The stimuli were then pretested in two ratings and one cloze pre-test with 75 participants in total. All pre-tests were conducted on participants from the undergraduate student pool for partial course credit. Thus, the demographic used in the pre-tests and the demographic used in the two actual experiments were very similar, which was important to ensure that the true/false/unknown conditions would work as intended.

Rating pre-test #1. Thirty-five native speakers of Canadian English (mean age = 22.0; 21 female, 12 male, 1 non-binary, 1 preferred not to report their gender) participated in the first rating pre-test. All participants reported to have lived in Canada most of their life, 30[85.7%] reported to have lived in Alberta most of their life. Their task was to rate the sentences using one of the five buttons ("definitely false", "maybe false", "I don't know", "maybe true", "definitely true").

For each sentence, we calculated the percentage of each of the five responses and evaluated how good the sentence fits its condition based on those responses. For true and false sentences, the percentage of "definitely true/false" ratings per sentence determined their score: those that got over 85% were rated as "perfect", 75-84% were rated as "excellent", 60-74% were rated as "very good", and 50-59% were rated as "good". True and false sentences that got less than 50% in "definitely true/false" but over 75% in "definitely true/false + maybe true/false" combined were rated as "acceptable", and those with 60-74% in those two categories were rated as "borderline". Everything else was rated as "poor". Sentences with an unknown truth value were evaluated as follows: "perfect" if the percentage of "I don't know" responses was 85% or higher, "excellent" if it was 75-84%, very good if it was 60-74%, good if it was 50-64%. The sentence was rated as "acceptable" if the percentage of "I don't know" didn't reach 50% but was still higher than any other rating choice. Everything else was rated as "poor." All poor sentences and most of the acceptable and borderline sentences were replaced; new sentences were then re-tested in the second pre-test.

Cloze pre-test. In addition to rating pre-tests, we also conducted a cloze probability pre-test. Twenty-three native speakers of Canadian English (15 female, 1 non-binary, 1 preferred not to report their gender, mean age = 19.0) completed a cloze pre-test. All participants reported to have lived in Canada most of their life, 18[78.3%] reported to have lived in Alberta most of their life. The task was to complete the sentences using one word. The results of the cloze pre-test were then

used to modify the sentences that got the rating of "poor", "acceptable", or "borderline" in the first rating pre-test before they were pre-tested again.

Rating pre-test #2. Seventeen participants participated in the pre-test. Two were excluded for having lived most of their life outside Canada. Thus, 15 participants entered the final analysis (mean age = 20.7; 13 female, 2 male). Fourteen (93.3%) reported having lived in Alberta most of their life. The sentences were then rated using the same procedure as in the first pre-test.

Based on the results of the pre-tests, the best 90 triplets were chosen. The characteristics of the final set of stimuli are shown in Table 5.2.

Table 5.2. The combined results of the rating pre-tests. Cells show the percentage of people who chose a specific label on the scale for each experimental condition. Standard deviations are in parentheses.

Condition	Mean pre-test rating (SD)						
	definitely true	**maybe true**	**definitely + maybe true**	**I don't know**	**maybe false**	**definitely false**	**definitely + maybe false**
True	71% (.22)	17% (.17)	88% (.12)	7% (.09)	2% (.04)	2% (.04)	4% (.06)
False	1% (.02)	4% (.06)	4% (.06)	9% (.11)	9% (.09)	78% (.17)	87% (.12)
Unknown	3% (.05)	21% (.11)	24% (.12)	62% (.15)	10% (.09)	4% (.06)	14% (.11)

Ninety filler sentences were used to distract the participants from the goal of the experiment. All fillers were grammatically and factually correct sentences containing very easy common knowledge information unrelated to Canada (e.g., "Bees collect nectar to make honey").

Initially, we planned to have twelve conditions in a fully crossed design: Truth Value (True, False, Unknown) x Accent (Native, Foreign) x Group (Canadian, Immigrant). However, presenting a speaker with a native accent as an Immigrant, particularly alongside speakers with a foreign accent presented as Canadian, was deemed to be highly unnatural. We thus decided to exclude this speaker condition (Immigrant native-accented). Thus, the final stimuli had nine conditions: Truth Value (True, False, Unknown) x Speaker (Canadian native-accented, Canadian foreign-accented, and Immigrant foreign-accented).

Nine lists were created by item rotation. Each participant saw only one sentence from each triplet in one of the Speaker conditions. The presentation of stimuli was pseudorandomized on the fly for each participant so that no more than three consecutive presentations of the same Speaker condition

and no more than three consecutive presentations of the same Truth Value condition were allowed. In total, each participant saw 180 sentences (90 experimental and 90 fillers).

5.2.2.2 Speakers

All speakers for this study were young women. Two native speakers of Canadian English were born and raised in Alberta. Native voices were always used in the Canadian native condition and were counterbalanced between participants. Foreign-accented speakers were both of European descent (one Italian and one Greek). As accent heaviness and intelligibility are partially independent (Munro & Derwing, 1995), we tried to recruit speakers whose accent was heavy but highly intelligible (pre-experiment accent ratings are discussed in the Results sections and agree with our assessments). Each speaker recorded a short introduction to present themselves that was used in the main experiment. The introductions were about 45 words long each and contained information about the speaker's name, birthplace, current occupation, and hobbies. Except for the birthplace and vague length of residency in Canada, all other details about the speakers were counterbalanced, and each speaker recorded all the versions of the introductions. All the names selected for the experiment could as easily be Italian/Greek as they could Canadian (Emma, Camilla, Sofia). The introductions are also available on OSF.

5.2.3 Procedure

The experiment was programmed in PsychoPy 3 software (Peirce et al., 2019). The procedure was similar to that of Foucart et al. (2019), as EEG activity was simultaneously recorded (but is not reported here). The participants were seated in a dimly lit, soundproof booth. They read sentences on a computer screen presented word by word. The experiment started with a short practice session during which the participants had to rate five short English sentences (e.g., "Giraffes have very long necks") as true or false on the same scale that was used in the main experiment. After the practice session, the experiment proper started.

To justify presenting the statements together with the picture of the speaker and their citizenship flag, we created a small cover story. The participants were told that the main goal of our research project is to collect a database of various trivia. We emphasized that we are trying to capture the linguistic and ethnic diversity of Canada, so the participants would hear statements from people from different backgrounds and from different parts of the country. The participants were told that

their task was to help us assess the veracity of their statements. They were then told that they would now hear the introductions from the three speakers randomly selected for them. During the introduction, the photo of the speaker, their name, group membership (Canadian/Immigrant) and the appropriate flag were shown on the screen. The three photos were taken from the Chicago Face Database (Ma et al., 2015) and featured white young women. The photos were closely matched in attractiveness (M = 4.7, SD = 0.08), age (M = 25.7, SD = 1.4), and femininity (M = 5.5, SD = 0.09). They were counterbalanced between participants. After each introduction, the participants had to rate on a 9-point scale how similar the speaker's accent was to their own and how foreign the speaker sounded. After the three introductions, the main trials began.

Each trial started with a fixation cross followed by the speaker's photo together with their name and their citizenship flag (Canadian/EU[16]). The group membership (Canadian/Immigrant) was not shown in the actual trials to avoid being too overt. The sentence was presented under the photo word by word in a white sentence-case Tahoma font on a dark gray background. After reading the sentence, the participants rated the truth-value of the sentence on a 5-point continuous scale (-2 = "Definitely false", -1 = "Maybe false", 0 = "I don't know", 1 = "Maybe true", 2 = "Definitely true"). The participants saw the labels of the scale but not the values. They had up to 7 seconds to respond. The sentence reading phase lasted for about one hour. The participants had four breaks, during which they were encouraged to rest and stretch.

After the main task, the participants completed a language background questionnaire[17] (available in Appendix B.7) and an IAT test, which is described below.

Implicit Association Test (IAT). We designed an IAT test largely based on Roessel et al. (2018) and Lee (2015). As target concepts, we used "CANADIAN ENGLISH" and "FOREIGN ACCENT", Both of these terms refer exclusively to language and thus cannot be confused with nationality, which helps avoid potential confounds (K. A. Lane et al., 2007). The attribute concepts were "GOOD" and "BAD". Attribute words were selected to represent the dimension of

[16] The EU flag was chosen to maintain consistency between the two foreign-accented speakers (since they had different countries of origin and would thus require two different flags when presented in the Immigrant condition). We did not deem it to matter, as long as it was not the Canadian flag.

[17] The participants also completed some other surveys (e.g., the political conservatism questionnaire), however, due to poor internal reliability of those surveys ($0.5 < \alpha < 0.65$) or massively skewed distributions, we did not analyze them and are not reporting them here.

knowledgeability: COMPETENT/INCOMPETENT, KNOWLEDGEABLE/IGNORANT, INTELLIGENT/DULL, EDUCATED/UNEDUCATED, SMART/STUPID. Attribute words did not differ significantly between the two categories either in frequency or in length according to Student's *t*-tests (all *p*s > .5). In line with common practice, the test consisted of seven blocks (Fig. 5.1).

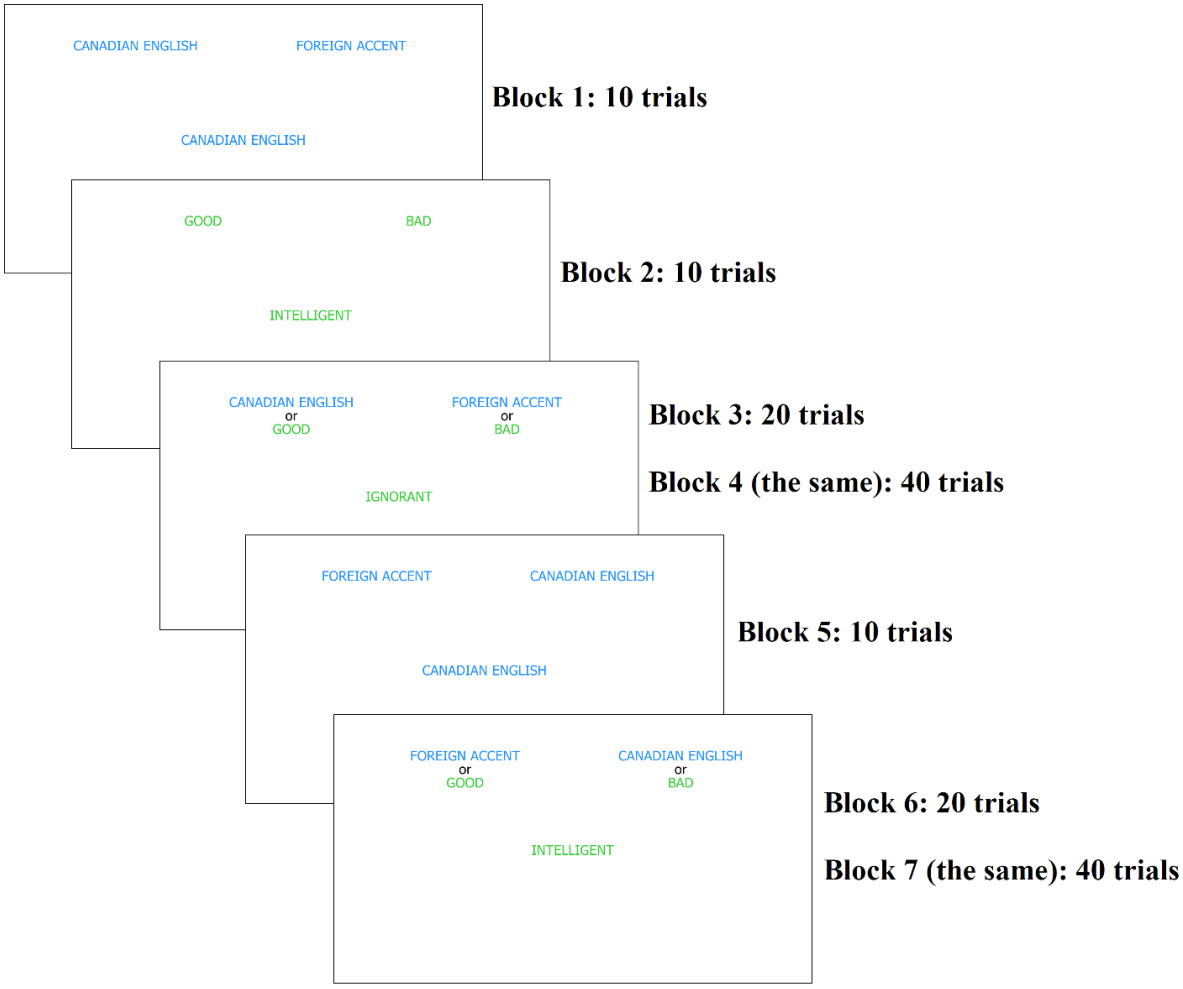

Figure 5.1. The structure of the IAT test.

The participants were instructed to sort the terms that appeared in the center of the screen into the categories in the top corners of the screen by pressing the 'A' or 'L' keys. In case of an incorrect response, a red cross appeared below the stimulus and stayed until a correct answer was given. The participants were instructed to go as quickly as possible while being accurate. Target words with their categories were presented before the target-practice block (Block 1) and attribute words with their categories were presented before the attribute-practice block (Block 2). The order of the

Blocks 3-4 and 6-7 was counterbalanced between participants so that half of them got the compatible block first (CANADIAN ENGLISH OR GOOD/FOREIGN ACCENT OR BAD) and the other half got the incompatible block first (FOREIGN ACCENT OR GOOD/CANADIAN ENGLISH OR BAD). Only the starting positions for categories in Block 1 were counterbalanced so that the initial combined block was also counterbalanced between subjects. The attribute block (Block 2) was not counterbalanced.

Before submitting their data, the participants were fully orally debriefed and provided with the opportunity to withdraw their data if they wanted to with no consequences and without losing their incentive. Nobody expressed a wish to withdraw their data, so all the participants could be retained.

5.2.4 Data analysis

All analyses were done in R (4.3.2) (R Core Team, 2020). R scripts together with the raw data are available on OSF. We used generalized additive mixed effects models of the Gaussian family with the *gam()* function from the *mgcv* (1.9–0) package (Wood, 2017) for the main analysis as they allow for the inclusion of complex random structures without convergence issues and facilitate the examination of non-linear effects. Post hoc comparisons were performed using the *emmeans* package with Tukey's honest significance test (1.8.9) (Lenth et al., 2023); effect sizes were calculated with the same package. The results were plotted with the *mgcViz* (0.1.11) (Fasiolo et al., 2023) and *ggplot2* (3.4.2) (Wickham et al., 2023) packages. For the IAT test, Cohen's *d* was obtained using the *cohens_d()* function from the *rstatix* package.

All models had random smooths for subject and item. Subject-specific random smooths were fitted for trial and both fixed factors in the model (Speaker, Truth Value). We also added a smooth for Truth Value by sentence pre-test rating, which significantly improved both the model's fit and coefficient of determination. To investigate the effect of the implicit prejudice, we compared the model with the IAT scores to the baseline model with the manipulated variables only using the *compareML()* function from the *itsadug* package (2.4.1) (van Rij et al., 2020). The *compareML()* function outputs a chi-square test of REML scores and an AIC difference between two models. We also used the *itsadug* package (2.4.1) and the *plot_diff()* function to examine the windows of significant differences between the conditions if they were significant. Since to date the only way to examine the effects of continuous predictors in interaction with several factorial predictors in a

GAMM model is to combine all factorial predictors into a grouping predictor with *n* levels (see (van Rij et al., 2020), we combined Speaker and Truth Value into a 9-level predictor Condition for examining the IAT scores. We started with the default *k* and then adjusted it depending on the results of *gam.check()*.

For truth rating analysis, we removed ratings < -1 (from "definitely false" before "maybe false", "maybe false" itself was not removed) for true sentences and all ratings > 1 (from "definitely true" before "maybe true", "maybe true" itself was not removed) for false sentences (1.3% of the data in total) as obvious errors in line with Lorenzoni et al. (2022).

IAT. An updated scoring algorithm by Greenwald et al. (2003) was used (D-scores). According to the algorithm, there are at least three different ways to calculate error penalty: 1) error trials = block mean of correct trials + 600 ms 2) error trials = block mean of correct trials + 2 SD 3) error trials = built-in error penalty (when participants have to correct their error themselves; latency to correct response is used). Since our participants were shown a red cross following an error and had to correct their response, we used the third method[18].

5.2.5 Results

5.2.3.1 IAT

One half of the participants (N = 15) had the congruent block first and the other half (N = 15) the incongruent block first. There were no data points exceeding 10000 ms and none of the participants had more than 10% of responses below 300 ms (the highest was 1.7%), so all data could be retained. Mean IAT score (where zero is the absence of bias in favor of any accent and scores above zero indicate a bias against foreign accent) was 0.29 (SD = 0.33, range -0.3 to 1.2). This agrees with the results of Roessel et al. (2018) who obtained a similar value for the dimension of competence in German. A one-sample *t*-test on IAT scores against zero showed that the bias against foreign accents was statistically significant ($t = 4.74$, df = 29, $p < 0.001$). Cohen's *d* was

[18] We additionally ran an analysis where we used the first method to see if that changes the result. There were virtually no differences between the two. When method #1 was used, mean IAT score was 0.31 (SD = 0.34, range -0.3 to 1.3). A one-sample *t*-test on IAT scores against zero showed that the bias against foreign accents was statistically significant in both analyses ($t > 5$, $p < 0.001$). There were differences in block effects. A simple linear model of IAT scores versus block order (congruent first vs. incongruent first) showed a significant effect ($t = -2.12$, $p = 0.041$) with the first method. IAT scores were larger in the compatible first versus incompatible first block order. We still decided to use the natural error penalty that was part of the IAT test design.

0.87, showing a large effect size. A simple linear model of IAT scores versus block order (congruent first vs. incongruent first) showed no significant effect of block order (t = -1.44, p = 0.161), although mean IAT scores were larger in the compatible first versus incompatible first block order (M = 0.38 vs. 0.20). This goes in line with previous research showing reduced bias scores when the incongruent block precedes the congruent one (Klauer & Mierke, 2005). The internal consistency of the IAT was determined via the Spearman-Brown split-half formula by applying the scoring algorithm separately to two mutually exclusive subsets of the IAT's combined-task trials (Glashouwer et al., 2013; Greenwald et al., 2003; Roessel et al., 2018) split using a random seed and then computing the Spearman-Brown coefficient between the scores of the two halves. It was satisfying (0.74).

5.2.3.2 Pre-experiment speaker evaluations

After the introductions, participants rated the similarity of the speaker's accent to their own (1 = very similar to 9 = very different) and how foreign the speaker sounded (1 = not at all foreign to 9 = very foreign). Generalized additive mixed models were used to analyze these results, with Speaker as a fixed predictor, crossed random effects for participant and trial (i.e. the order in which the speakers were rated) and a random slope for participant and trial. Pairwise comparisons were performed to estimate all contrasts. The results of the evaluations are depicted in Fig. 5.2.

Slight but notable differences were observed between the two foreign-accented voices, particularly in the Canadian foreign condition. Voice #2 showed substantially more variation between the Immigrant foreign and Canadian foreign conditions in both accent similarity (8.0 vs. 6.9) and foreignness (6.5 vs. 6.0) ratings, compared to voice #1, which had similar ratings in both conditions (accent: 7.9 vs. 8.0; foreignness: 7.3 vs. 7.2). Student t-tests between the two conditions showed that this difference was marginally significant for voice #2 for the accent rating (t = -2.01, p = 0.051) but not the foreignness rating. There are two possible explanations for this finding, not mutually exclusive. First, the foreign accent of voice #2 was milder than that of voice #1. It might be that there is a threshold upon which the group manipulation becomes effective. Alternatively, if the accent is really strong (albeit highly intelligible) the participants may question if the speaker has indeed lived in Canada for long enough to become a (New) Canadian, despite the fact that scientifically, accent strength does not always correlate with length of residence. This issue warrants further investigation.

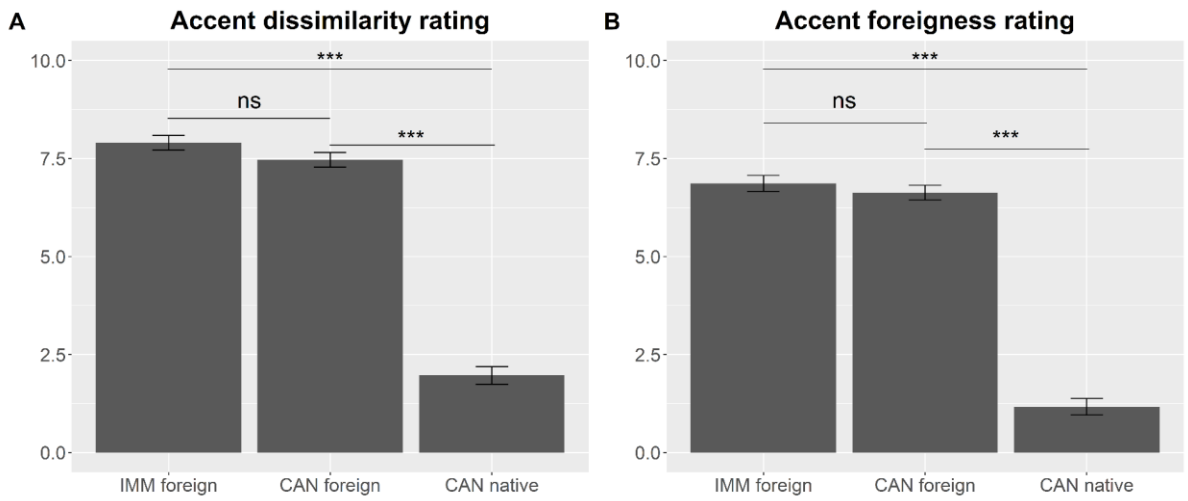

Figure 5.2. Barplots of pre-experiment speaker evaluations in Experiment 1: accent dissimilarity (A) and accent foreignness (B). Error bars indicate standard error of the mean. Asterisks indicate statistical significance ('***' $p < 0.001$, '**' $p < 0.01$, p '*' < 0.05, '.' $p < 0.06$, ns = non-significant).

5.2.3.3 Truth ratings

Mean ratings for each condition are reported in Table 5.3. As can be seen from the table, unknown sentences by foreign-accented speakers were rated lower for veracity than those by the Canadian native speaker.

Table 5.3. Mean truth ratings for all experimental conditions in Experiment 1 (SDs in parentheses).

Speaker	Truth Value		
	True	False	Unknown
Canadian native	1.53 (0.7)	-1.66 (0.7)	0.21 (0.7)
Canadian foreign	1.61 (0.6)	-1.61 (0.7)	0.10 (0.6)
Immigrant foreign	1.59 (0.7)	-1.61 (0.7)	0.10 (0.7)

To investigate these results further, we used generalized additive mixed modelling. For ease of interpretation, we contrast-coded the Speaker condition and sum-coded the Truth Value condition. This approach was chosen to enhance the visibility of speaker contrasts specifically for unknown sentences, while at the same time avoiding the complexity of interpreting interactions across multiple levels (e.g., speaker1 vs. speaker2 × truth value1 vs. truth value2). The results of the main GAMM model are reported in Table 5.4.

Table 5.4. Summary of the GAMM model for Experiment 1 with manipulated variables only. Model's formula: truth rating ~ speaker*truth value + s(participant, bs = 're') + s(triplet number, bs = 're') + s(participant, trial, bs = 're') + s(participant, speaker, bs = 're') + s(participant, truth value, bs = 're') + s(truth value, pretest rating, bs = 're'). Asterisks indicate statistical significance (*p < 0.05, **p < 0.01 and ***p < 0.001). Significant effects are in bold. Deviance explained = 83.1%.

Predictors	***Estimates***	***CI***	***p***	
Intercept (grand mean)	0.02	-0.11 – 0.15	0.752	
CAN foreign vs. CAN native	0.01	-0.05 – 0.07	0.765	
IMM foreign vs. CAN native	-0.01	-0.07 – 0.05	0.821	
unknown	0.15	-0.04 – 0.33	0.119	
false	**-1.55**	**-1.74 – -1.36**	**<0.001**	***
CAN foreign vs. CAN native × unknown	**-0.12**	**-0.21 – -0.04**	**0.003**	**
IMM foreign vs. CAN native × unknown	**-0.10**	**-0.18 – -0.02**	**0.017**	*
CAN foreign vs. CAN native × false	0.04	-0.04 – 0.12	0.326	
IMM foreign vs. CAN native × false	0.05	-0.03 – 0.13	0.237	
Smooth term (participant)			0.931	
Smooth term (triplet number)			**0.040**	*
Smooth term (participant, trial)			0.450	
Smooth term (participant, speaker)			0.375	
Smooth term (participant, truth value)			**<0.001**	***
Smooth term (truth value, pretest rating)			**<0.001**	***

As can be seen from the model, false sentences were rated significantly lower than the grand mean but unknown sentences did not differ from the grand mean. In addition, both two-way interactions for unknown sentences were significant, indicating that statements by both foreign-accented speakers were rated lower than those by the Canadian native speaker. Planned post hoc comparisons using the Tukey correction showed that the difference between Canadian native and Canadian foreign speaker for unknown sentences was marginally significant (*Est.* = 0.12, $t = 2.26$, $p = 0.061$, effect size = 0.19), same as the difference between Canadian native and Immigrant foreign speakers (*Est.* = 0.11, $t = 2.08$, $p = 0.094$, effect size = 0.17). The difference between the two foreign speakers was not significant (*Est.* = -0.01, $t = -0.18$, $p = 0.982$, effect size = -0.02). All other differences between speakers in true and false conditions were also not significant (ps > 0.1).

IAT scores did not improve the model's fit.

5.3 Discussion

We found some evidence that foreign-accented speakers may be indeed trusted less on domain-specific unknown statements, albeit very weak. The main model showed that such statements attributed to both foreign-accented speakers (regardless of their group membership) were rated as less true than the same statements attributed to a native-accented speaker, although in the post hoc comparisons these differences were only marginally significant.

In light of these results, it is important to reiterate that the participants only heard the accent during the introductory phase before the main experiment. During each trial, all they saw on the screen was the photo of the person who purportedly "came up with this fact", their name (all of which are common in Canada), and the flag indicating the speaker's citizenship (Canada or the European Union). Hence, any effect of speaker accent in this study was due to the fact that participants actually *tracked* this feature even when the person in question was not even speaking (since both foreign-accented speakers were rated significantly lower than the native speaker, despite one of them being presented with a Canadian flag).

We did not find any evidence that individual differences in implicit attitudes toward foreign-accented speakers affected the results. However, importantly, we did find a very strong bias against foreign-accented speakers in the IAT using knowledgeability attributes: mean slowdown in the incongruent condition (foreign accent paired with positive attributes and native accent paired with negative attributes) was 290 milliseconds, and the effect size was large. Thus, at least subconsciously, foreign-accented speakers *were* considered less knowledgeable. This is good evidence that the effect found by Lev-Ari & Keysar (2010) was real, although the results of the IAT do not fit with the authors' explanation regarding processing fluency. Indeed, if native speakers generally consider non-native speakers less knowledgeable, it is not the disrupted processing fluency per se that leads to lower truth ratings, it is the *attitudes* toward non-native speakers. This also explains why disrupting fluency in other ways (e.g., using white noise) does not lead to lower truth ratings (Souza, 2012), and why this experiment found lower truth ratings for unknown statements attributed to non-native speakers even without any oral speech (although this evidence was weak, as explained above).

One of the reasons for weak results in this study could be the method used. Even if there is domain-specific mistrust toward foreign-accented speakers which does not boil down to processing difficulty, it is arguable that the photo of the speaker and their national flag might be too easy to overlook, allowing participants to focus mainly on the statement. Experiment 2 thus tested the exact same sentences in the auditory modality.

In addition, given that previous studies reported an effect of listeners' ratings of non-native speakers on their auditory speech perception (Ingvalson et al., 2017), we decided to add explicit evaluations of the speakers on the status and solidarity dimensions to investigate whether those ratings modulated participants' truth ratings.

Experiment 2

5.4 Materials and methods

5.4.1 Participants

Thirty-six self-identified native English speakers with no history of neurological or psychiatric conditions participated in the experiment. All participants were students and received 20 CAD per hour of their participation (N = 35) or partial course credit (N = 1). Same as in Experiment 1, data from participants who have not lived in Alberta most of their life was not analyzed (N = 3). The final sample thus included 33 participants (mean age = 21.6, range: 18–29, SD: 2.8). Twenty-five participants (76%) self-identified as female, the rest self-identified as male. Fourteen participants (42%) were bilingual (defined as native-like in two languages). All participants self-identified as Canadian.

5.4.2 Materials

The materials were the same as in Experiment 1, with the only difference that the sentences were auditory. Each speaker (who also recorded the introductions used in Experiment 1) recorded sentences in a fully randomized order (i.e. true, false, and unknown sentences from the same triplet were not recorded sequentially to minimize prosodic confounds).

5.4.2.1 Acoustic analysis

All stimuli were scaled to 70dB. Stimuli descriptive statistics such as mean duration, mean pitch, mean speech rate (the number of syllables per time in seconds; de Jong & Wempe, 2009) are reported in Table 5.5.

Table 5.5. The comparison of acoustic features between the conditions. SDs are reported in parentheses.

Condition		Prosodic parameters		
Truth Value	**Accent**	**Mean duration, ms**	**Mean pitch, Hz**	**Mean speech rate, nsyll/s**
True	Native	2657 (471)	215 (15)	4.9 (0.6)
	Foreign	3382 (672)	226 (27)	3.9 (0.5)
False	Native	2666 (473)	212 (15)	4.8 (0.6)
	Foreign	3411 (706)	228 (27)	3.9 (0.5)
Unknown	Native	2804 (473)	209 (16)	4.8 (0.7)
	Foreign	3583 (706)	227 (29)	3.9 (0.5)

Three GAMM models were then fitted to assess the significance of differences between the conditions, one per each acoustic measure. Each model had Accent × Truth Value as fixed factors (dummy-coded) and a random smooth for item and voice (native #1, native #2, foreign #1, foreign #2). There was a main effect of Accent for all the inspected acoustic measures—foreign-accented sentences were longer, higher in pitch, and slower in speech rate (*p*s < 0.001). In addition, several interactions were significant. Full model outputs are available in Appendix C.3. Table 5.6 shows the results of pairwise comparisons for Accent × Truth Value with Tukey correction.

Table 5.6. Pairwise comparisons of acoustic features between the conditions. Significant results are in bold.

Contrast	Duration		Pitch		Speech rate	
	t	***p***	***t***	***p***	***t***	***p***
True: native - foreign	**-12.05**	**<0.001**	**-6.67**	**<0.001**	**16.10**	**<0.001**
False: native - foreign	**-12.38**	**<0.001**	**-8.99**	**<0.001**	**14.91**	**<0.001**
Unknown: native - foreign	**-12.95**	**<0.001**	**-10.01**	**<0.001**	**16.03**	**<0.001**
Native: unknown - false	2.29	0.199	-1.67	0.553	0.17	1.000
Native: unknown - true	2.44	0.144	**-3.20**	**0.018**	-0.43	0.998
Native: false - true	0.15	1.000	-1.54	0.642	-0.59	0.992
Foreign: unknown - false	**2.87**	**0.048**	-0.65	0.987	-0.95	0.932
Foreign: unknown - true	**3.35**	**0.011**	0.14	1.000	-0.35	0.999
Foreign: false - true	0.48	0.997	0.78	0.970	0.60	0.991

As can be seen from the table, foreign-accented unknown sentences were marginally longer than false sentences and significantly longer than true sentences. Native-accented unknown sentences were lower in pitch than true sentences. The rest was not significant.

5.4.3 Procedure

The procedure was the same as in Experiment 1 except for the following differences: 1) The sentences were presented orally rather than in the written form; 2) A 5-point continuous scale used in the previous experiment was replaced by a 100-point ordinal scale to allow for more granularity (note that such wide scale is not uncommon in this line of research—Lorenzoni et al., 2022 used a 100-point ordinal scale, whereas Barlow et al., 2024 and Boduch-Grabka & Lev-Ari, 2021 used a 100-point continuous scale); 3) Another pre-experiment speaker evaluation question was added after the introductions, "How well did you understand the speaker?" (0 = understood nothing to 9 = understood everything); 4) Three post-experiment speaker evaluation questions were added right after the main task, one to cover the dimension of status ("How knowledgeable was the speaker?") and two more to cover the dimension of solidarity ("How do you feel about the speaker overall?", "How pleasant was the speaker?"). Importantly, only the photo and the name of the speaker were shown during the evaluation stage (i.e. the flag indicating group membership was *not* shown); 5) Four more questions were added to the language background questionnaire (see Appendix B.7)—in particular, we collected information on participants' bilingualism since it has been suggested that monolingual speakers of standard dialects may more readily categorize foreign-accented speakers as an outgroup than bilingual speakers or those who speak a non-standard language variety (Ryan, 1983). In addition, as argued by the authors, experience with other language varieties and social groups can also increase or decrease accent tolerance as it increases the sensitivity to the social consequences of being a speaker of a particular language variety.

5.4.4 Data analysis

The response variable was centered. The rest of the analysis was done in the same way as in Experiment 1. Bilingualism was analyzed as a fixed term (binary predictor) rather than a smooth. R scripts together with the raw data are available on OSF.

The rating scale was 101-point ordinal (0 = “definitely false” and 100 = “definitely true”). The participants only saw the labels and the numbers for the ends of the scale, but nothing in-between. To remove incorrect answers, we applied the same cut-off point as for Experiment 1 (which in this case would be 25 for true and 75 for false sentences) and that produced a very similar data loss (1.6%).

5.4.5 Results

5.4.5.1 IAT

Eighteen participants had the congruent block first and 15 the incongruent block first. There were no data points exceeding 10 000 ms and none of the participants had more than 10% of responses below 300 ms (the highest was 2.5%), so all data could be retained. Mean IAT score was 0.39 (SD = 0.3, range -0.2 to 1.2), much higher than in Experiment 1. A one-sample *t*-test on IAT scores against zero showed that the bias against foreign accents was statistically significant ($t = 6.77$, df = 32, $p < 0.001$). Cohen’s *d* was 1.18, showing a very large effect size, larger than in Experiment 1 (0.87). A linear model of IAT scores versus block order (congruent first vs. incongruent first) again showed no effect of block order ($t = -1.46$, $p = 0.2$). Descriptively, IAT scores were still a bit larger in the compatible first versus incompatible first block order (M = 0.47 vs. 0.30). The Spearman-Brown split-half coefficient was good (0.81).

5.4.5.2 Pre-experiment speaker evaluations

Accent dissimilarity and foreignness were assessed with the same questions and scales as in Experiment 1. We additionally asked participants to rate the intelligibility of each speaker (“How well did you understand the speaker?” where 1 = “understood nothing” and 9 = “understood everything”). The same statistical analyses were conducted. The results of the evaluations are depicted in Fig. 5.3. Although the Canadian foreign speaker was rated as less intelligible than the Canadian native speaker (M = 8.6, SD = 0.7 vs. M = 9.0, SD = 0.0), it is worth noting that intelligibility ratings were at ceiling on a 9-point scale, so any effects we found in the main experiment are extremely unlikely to be due to comprehension issues. The Canadian native speaker was consistently rated as absolutely intelligible by every participant.

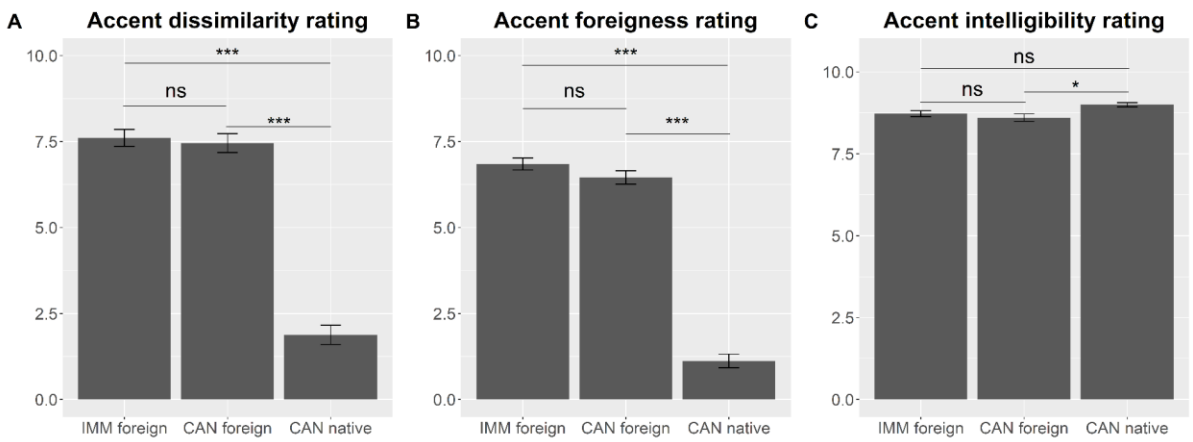

Figure 5.3. Barplots of pre-experiment speaker evaluations in Experiment 1: accent dissimilarity (A), accent foreignness (B), and accent intelligibility (C). Error bars indicate standard error of the mean. Asterisks indicate statistical significance ('***' $p < 0.001$, '**' $p < 0.01$, p '*' < 0.05, '.' $p < 0.06$, ns = non-significant).

Group manipulation seemed ineffective for accentedness rating for voice #1, whereas voice #2 showed a 0.4 reduction in the Canadian foreign condition. Although the *t*-tests were insignificant (*p*s > 0.5), this result is fully in line with Experiment 1 and thus warrants further investigation with larger samples. The group manipulation reduced perceived foreignness by 0.3 for voice #1 and 0.5 for voice #2, though *t*-tests were insignificant (*p*s > 0.4). Intelligibility for both voices was unaffected by the group manipulation (*p*s > 0.4).

5.4.5.3 Post-experiment speaker evaluations

Participants rated the speakers on three attributes using a 100-point scale: their overall impression about the speaker (0 = "very negative" to 100 = "very positive"), how knowledgeable the speaker was (0 = "not at all knowledgeable" to 100 = "very knowledgeable") and how pleasant the speaker was (0 = "very unpleasant" and 100 = "very pleasant"). The same models were fit as for the pre-experiment ratings. The results are shown in Fig. 5.4. In summary, the native speaker was consistently rated higher than both non-native speakers on every dimension.

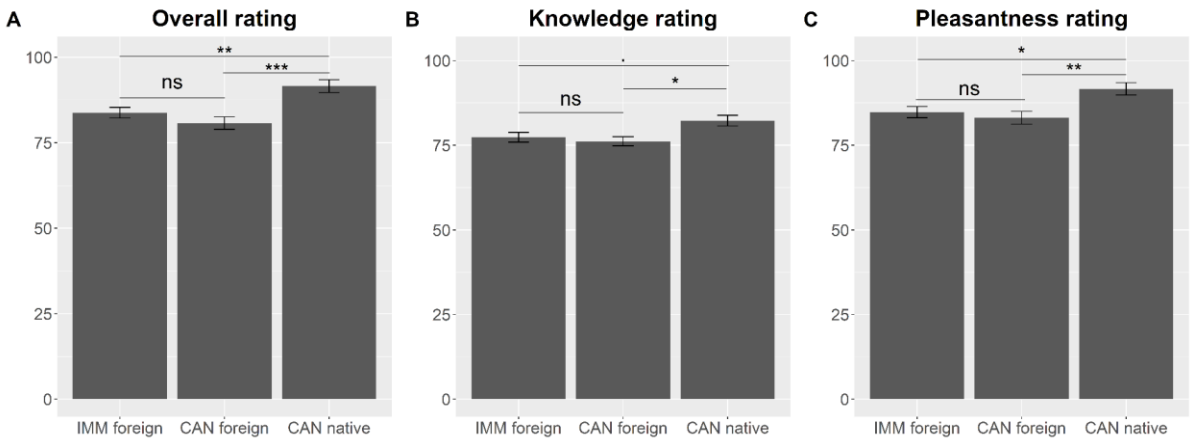

Figure 5.4. Barplots of post-experiment speaker evaluations: overall speaker rating (A), speaker knowledge rating (B), speaker pleasantness rating (C). Error bars indicate standard error of the mean. Asterisks indicate statistical significance ('***' $p < 0.001$, '**' $p < 0.01$, p '*' < 0.05, '.' $p < 0.06$, ns = non-significant).

Importantly, we decided to examine whether knowledgeability ratings in the two foreign-accented speaker conditions were correlated with IAT scores (i.e. whether explicit attitudes were correlated with generalized implicit attitudes). The results showed that they indeed were correlated, albeit weakly: $r = 0.15$, $t = 4.71$, $p < 0.001$ for the Canadian native speaker, $r = 0.12$, $t = 3.75$, $p < 0.001$ for the Canadian foreign speaker, and $r = -0.07$, $t = -2.30$, $p = 0.022$ for the Immigrant foreign speaker.

5.4.5.4 Truth ratings

Mean ratings for each condition are reported in Table 5.7. It appears from the descriptive results that all statements by the Canadian foreign-accented speaker were rated higher for veracity, regardless of their inherent truth value.

Table 5.7. Mean truth ratings for all experimental conditions in Experiment 2 (SDs in parentheses).

Speaker	Truth Value		
	True	False	Unknown
Canadian native	40.58 (15.2)	-41.45 (16.3)	-0.58 (19.8)
Canadian foreign	40.86 (16.0)	-39.98 (17.4)	2.42 (20.3)
Immigrant foreign	38.88 (17.3)	-41.15 (16.3)	0.05 (20.2)

To investigate these results further, we used generalized additive mixed modelling. Same as in Experiment 1, the speaker condition was contrast-coded while the truth value condition was sum-coded. The results of the main GAMM model are reported in Table 5.8.

Table 5.8. Summary of the GAMM model for Experiment 2 with manipulated variables only. Model's formula: truth_rating ~ speaker*truth_value + s(participant, bs = 're') + s(triplet_number, bs = 're') + s(participant, trial, bs = 're') + s(participant, speaker, bs = 're') + s(participant, truth_value, bs = 're') + s(truth_value, pretest_rating, bs = 're'). Asterisks indicate statistical significance (*p < 0.05, **p < 0.01 and ***p < 0.001). Significant effects are in bold. Deviance explained = 82%.

Predictors	***Estimates***	***CI***	***p***	
Intercept (grand mean)	0.35	-2.62 – 3.33	0.816	
CAN native vs. CAN foreign	**-1.57**	**-3.02 – -0.11**	**0.035**	*
IMM foreign vs. CAN foreign	**-1.81**	**-3.27 – -0.36**	**0.015**	*
unknown	0.80	-3.32 – 4.93	0.703	
false	**-37.95**	**-42.21 – -33.69**	**<0.001**	***
CAN native vs. CAN foreign × unknown	-1.51	-3.56 – 0.54	0.149	
IMM foreign vs. CAN foreign × unknown	-0.67	-2.72 – 1.39	0.525	
CAN native vs. CAN foreign × false	0.04	-2.03 – 2.10	0.972	
IMM foreign vs. CAN foreign × false	0.51	-1.54 – 2.57	0.625	
Smooth term (participant)			0.713	
Smooth term (triplet number)			**0.004**	**
Smooth term (participant, trial)			0.328	
Smooth term (participant, speaker)			0.830	
Smooth term (participant, truth value)			**<0.001**	***
Smooth term (truth value, pretest rating)			**<0.001**	***

The model confirmed a main effect of Speaker—the participants believed all statements by the Canadian foreign-accented speaker more than those by the Canadian native or the Immigrant foreign-accented speaker. No two-way interactions were significant; hence no post hoc comparisons were needed.

IAT. IAT scores significantly improved the baseline model ($p < 0.001$). Table 5.9 shows partial model summary for IAT smooths only. Fig. 5.5 shows difference plots for unknown sentences. As can be seen in panel A, people with scores above average (M = 0.39) rated unknown sentences lower when the foreign-accented speaker was presented as Immigrant compared to Canadian. Panel B shows that people with low IAT scores (< 0) rated unknown statements by a Canadian native speaker higher than those by a Canadian foreign-accented speaker, whereas people with

fairly high scores seem to be affected in the opposite way—although the slope is not straightforwardly downward but rather starts to slowly rise again. Panel C shows that people with low IAT scores (< 0) rated unknown statements by a Canadian native speaker higher than those by an Immigrant foreign-accented speaker. Overall, the general trend for the first and third plots is that the more prejudiced the participant is, the higher they rate speakers presented with a Canadian flag and the lower they rate the speaker presented with a European flag.

Table 5.9. Partial output of the GAMM model for Experiment 2 showing significance of each IAT smooth. Asterisks indicate statistical significance (*p < 0.05, **p < 0.01 and ***p < 0.001).

Predictors	***p***	
Smooth term (IAT score) × unknown CAN native	0.035	*
Smooth term (IAT score) × unknown CAN foreign	0.218	
Smooth term (IAT score) × unknown IMM foreign	0.006	**
Smooth term (IAT score) × false CAN native	0.215	
Smooth term (IAT score) × false CAN foreign	0.019	*
Smooth term (IAT score) × false IMM foreign	0.472	
Smooth term (IAT score) × true CAN native	0.668	
Smooth term (IAT score) × true CAN foreign	0.432	
Smooth term (IAT score) × true IMM foreign	0.431	

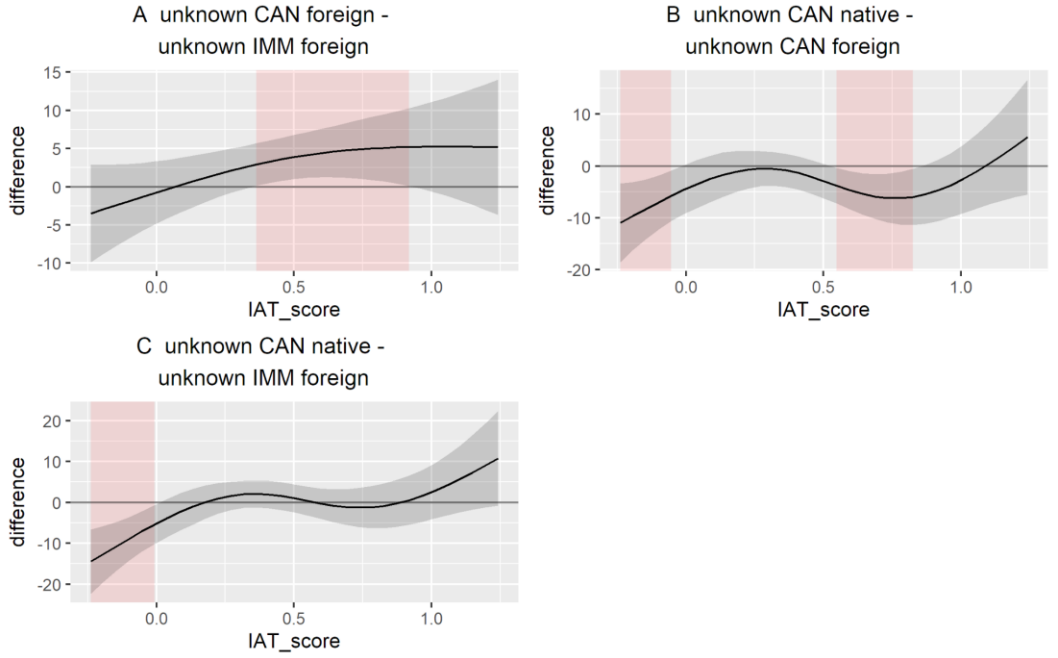

Figure 5.5. Difference plots smoothed by IAT test scores. Higher scores indicate greater bias against foreign accents and in favour of native accents. Y-axis shows an estimated difference in truth ratings. Shaded areas indicate windows of significant differences.

Bilingualism. Bilingualism significantly improved the baseline model ($p < 0.001$). There were two significant interactions: CAN native vs. CAN foreign × unknown × if bilingual [1] (*Est.* = 4.70, *CI* = 0.55 – 8.85, $p = 0.026$) and IMM foreign vs. CAN foreign × unknown × if bilingual [1] (*Est.* = 4.56, *CI* = 0.41 – 8.71, $p = 0.031$). Fig. 5.6 shows interaction plots. Monolingual participants rated unknown sentences by the Canadian foreign speaker higher than those by either Canadian native or Immigrant foreign speakers (i.e. the main effect found in this study), whereas bilingual participants did not exhibit this effect and rated unknown sentences the same regardless of the speaker. In short, bilingual participants were not affected by the group manipulation.

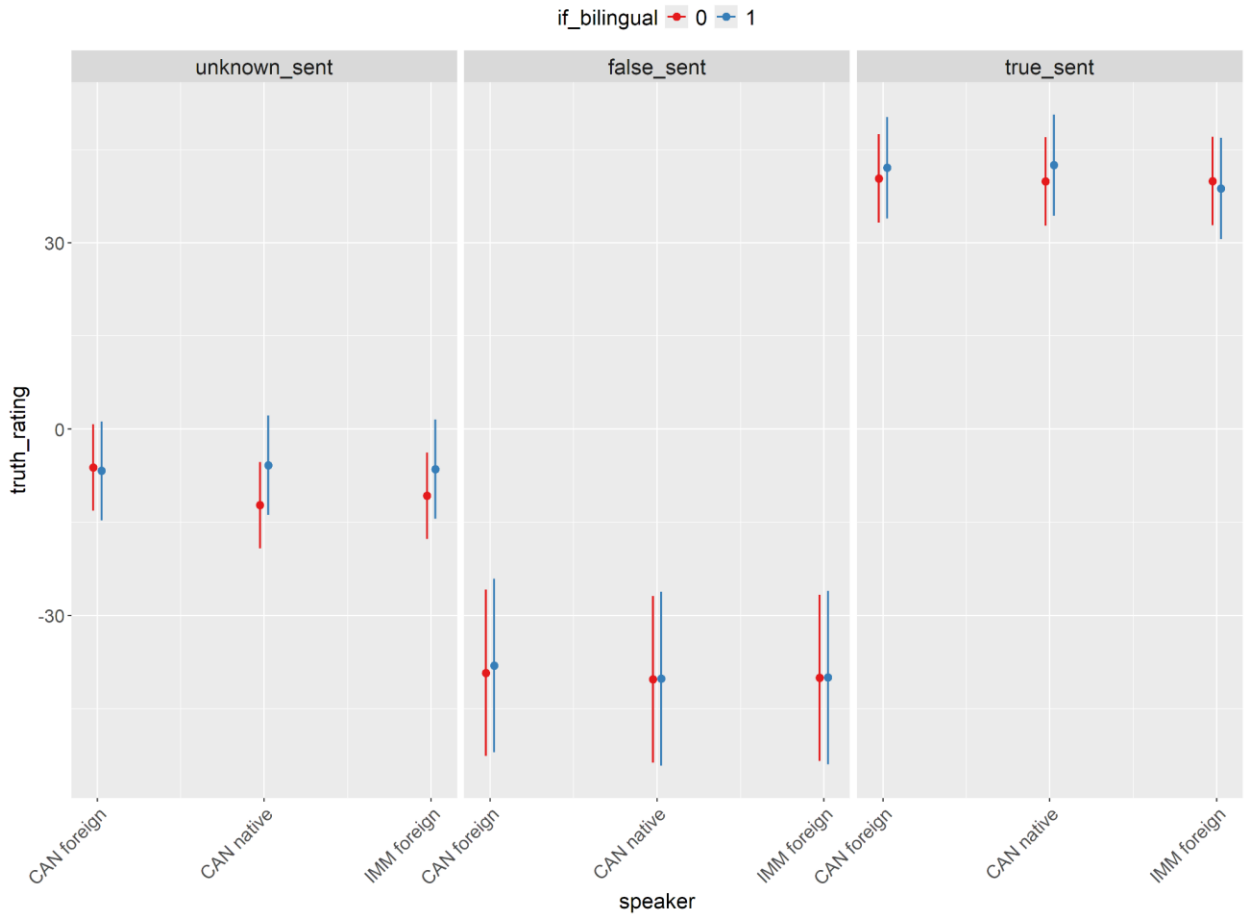

Figure 5.6. Predicted values (marginal effects) of the interaction between bilingualism, speaker, and truth value. The dependent variable is truth rating.

Because Ingvalson et al. (2017) found an effect of listeners' ratings of non-native speakers on their speech perception ability, we additionally examined whether the listeners' ratings of the speakers

modulated their truth ratings. Although knowledgeability ratings by themselves marginally improved the model's fit (p = 0.054) and the average speaker ratings across the three scales (knowledgeability, likeability, pleasantness) significantly improved the model's fit (p = 0.037), none of the individual smooths by condition were significant, hence we did not investigate the models further.

5.4.6 Discussion

There were several findings from this experiment. First and foremost, in contrast to Experiment 1 where domain-specific unknown statements by both foreign-accented speakers were rated as less true, *all* statements in Experiment 2 were rated higher for veracity when they came from the Canadian foreign-accented speaker. Why would that be the case? One plausible explanation emerges. Since all our critical statements were about Canada and the foreign-accented speaker was presented as Canadian, listeners might perceive them as someone who has put significant effort into learning about and integrating into Canadian culture. This perception could be influenced by the fact that obtaining a Canadian passport requires an individual to pass a citizenship test. The test consists of questions about Canada's history, geography, economy, government, laws, and symbols. Trivia statements that we used actually closely resemble the material of the test (the number of provinces, population, flags, etc.). Anecdotally, it is not uncommon for new Canadian citizens to have more detailed, up-to-date knowledge of some facts about Canadian history and geography compared to locals, who may have learned this information in school long ago. It is also noteworthy that, even though this difference was not significant, unknown statements by the Immigrant foreign speaker were rated higher on truth than those by the Canadian native speaker, with the latter being rated the lowest of all three. This directly contradicts the results of Experiment 1 and the intelligibility-based account in general, which would predict that native speakers are always rated as more credible due to better intelligibility. Even if we assume that presenting the foreign-accented speaker as Canadian improved accent processing (which would contradict the results of Calkins et al., 2023), it would still not improve intelligibility beyond that of a native speaker. The findings strongly suggest that credibility is not just about ease of comprehension but also about social perceptions and group dynamics. Listeners do not process accents in isolation but consider it alongside other information about the speaker's background and presumed knowledge. It also suggests that group membership change may be modality-specific. In general, this is not

surprising as modality-specific effects in language processing have been attested (see, for instance, Souza, 2012).

Perhaps even more surprising was a clear mismatch between actual truth judgments made in real-time and explicit evaluations of the speakers made retrospectively, specifically for the Canadian foreign-accented speaker. A similarly mismatching effect was observed in Stocker (2017): whereas participants rated different speech communities differently on credibility in explicit surveys, truth judgments did not differ between native and foreign-accented speakers. The reason for this discrepancy in our study is likely to be that the participants did not *actually* remember who said what and how knowledgeable the speakers really were in the moment. Practically speaking, every speaker produced the same number of true, false, and unknown utterances. However, it is quite obvious that the participants did track group membership—although *no flag or group label* were shown during the evaluation stage, the Canadian native speaker was still rated significantly higher on every single dimension than both foreign-accented speakers. These results are fully in line with several studies showing that speakers of non-standards accents are consistently evaluated lower on numerous dimensions, including status-related variables associated with knowledge and education (Fuertes et al., 2012).

Importantly, we did find an effect of implicit prejudice on truth ratings. The overall trend was that higher prejudice against foreign-accented speakers correlated with higher truth ratings of unknown sentences when they were produced by speakers introduced as Canadian. This effect was the most pronounced and straightforward for Immigrant foreign-accented speakers. Participants who scored above average on the IAT test rated unknown sentences much lower when the foreign-accented speaker was introduced as Immigrant vs. Canadian.

Finally, only monolingual participants showed a preference for the Canadian foreign speaker when rating unknown sentences for truth. Bilingual participants were unaffected by either accent or the group membership manipulation, rating unknown sentences equally regardless of the speaker. It is important to note that 42 percent of our participants were bilingual, which is not a small percentage. This difference between bilingual and monolingual participants merits further investigation. It could suggest that bilingual individuals, who have experience with multiple languages and potentially multiple cultural contexts, are less susceptible to accent-based or group membership-

based biases and are more attuned to the idea that accent is not necessarily an indicator of knowledge or credibility.

5.5 General discussion

The present research, consisting of two experiments in different modalities, aimed to investigate the effect of accentedness on perceived veracity of domain-specific statements and explore the underlying mechanisms of this phenomenon. Of particular interest to us was whether the "incredibility effect", if observed, could be diminished or overridden by changing the speaker's group membership—i.e. in this case, presenting foreign-accented speakers as Canadian citizens.

In short, neither study provided support for the fluency-based account. Experiment 1 showed weak evidence that unknown statements by non-native speakers may be perceived as less true. However, due to written stimuli, these effects cannot be explained by reduced intelligibility. Participants were exposed to the accent only during an introductory phase, suggesting that accent information was retained and applied in the absence of ongoing auditory input, thus lending support to outgroup-based accounts. The fact that both non-native speakers were rated lower than the native speaker implies that changing group membership was unsuccessful. One potential explanation for this outcome may be attentional. With written sentences only, participants may have focused predominantly on the most salient feature (presumably the accent heard *prior* to the experiment) while disregarding the visual cues (flags). Future research not constrained by simultaneous EEG data collection might benefit from employing a blocked design, allowing participants to concentrate on one speaker at a time, potentially amplifying speaker-related differences. It is crucial to note that the non-native "incredibility" effect found in Experiment 1 was only marginally significant in post hoc comparisons and of small magnitude (0.1 on a 5-point scale). Given mixed results observed in previous research as well as the social significance of this research agenda, it is paramount that the results observed in this study are validated with a larger participant sample.

Experiment 2 yielded surprising results that partly challenge our initial hypotheses and also contradict accounts based on processing fluency. Contrary to Experiment 1, the participants attributed higher veracity to all statements made by the Canadian foreign-accented speaker compared to those of the Canadian native or the Immigrant foreign-accented speaker. This

unanticipated finding suggests a potential "knowledgeable immigrant" effect, wherein listeners may perceive foreign-accented individuals who have obtained citizenship as having invested considerable effort in familiarizing themselves with the country and its culture. Although we expected group membership change to raise the credibility of foreign-accented speakers, we did not expect this effect to surpass the credibility ascribed to native speakers. The discrepancy between the results of Experiments 1 and 2 highlights the importance of methodology in accent research. The advantage observed for foreign-accented speakers labeled as Canadian in Experiment 2 implies that actual oral speech may be necessary to render group membership changes more effective, potential because accent combined with the group label serves as a constant perceptually salient reminder of one's effort to naturalise.

We also replicated the fundamental social psychology finding that non-native speakers consistently receive more negative social evaluations than native speakers on various dimensions (in our study: likeability, knowledgeability, and pleasantness). Crucially, during these post-experiment evaluations, only the photo of the speaker from the main experiment was shown. This suggests that: 1) the participants retained a robust memory of the speaker's accent, 2) explicit group membership information was not retained or utilized in these judgments. The lack of differentiation between Canadian and Immigrant foreign-accented speakers in post-experiment evaluations highlights the potential limitations of explicit group membership manipulations in altering deeply ingrained accent-based attitudes, at least for social evaluations.

One of the most significant findings of this study concerns implicit attitudes toward foreign accents. Despite the mixed results in explicit truth ratings, the study revealed a robust implicit bias against foreign-accented speakers in the knowledgeability domain as measured by the Implicit Association Test. Native Canadian speakers experienced a significant slowdown in response when they had to press the same button when "foreign accent" and positive knowledge attributes like "intelligent" and "competent" appeared on the screen compared to "Canadian English". They also experienced a slowdown when "Canadian English" and negative knowledge attributes like "ignorant" and "uneducated" were mapped onto the same button compared to "foreign accent". This automatic delay is generally interpreted as indicative of cognitive dissonance, revealing unconscious prejudices beyond volitional control. An effect of a similar magnitude was found in native German speakers when they had to sort "akzent" (explained as German with a non-native

accent) and positive competence attributes, as well as, albeit to a lesser extent, positive attributes on the dimensions of affect and trust (Roessel et al., 2018). Crucially, we demonstrated that the strength of this implicit prejudice modulates truth ratings. Contrary to our prediction, participants with stronger implicit biases showed more pronounced effects of group membership on truth ratings. Specifically, they ascribed lower truthfulness to statements when spoken by a foreign-accented speaker introduced as Immigrant compared to when the same speaker was introduced as Canadian. This effect was significant for people to the right of the mean (more prejudiced than average). To the best of our knowledge, this is the first time an IAT measuring generalized prejudice against foreign-accented speakers was done in Canada and the first time it was used predictively in psycholinguistics. It is particularly interesting that the IAT was only predictive of ratings in Experiment 2, when aural stimuli were used.

One important issue to address is, of course, to what extent this "knowledgeable immigrant" finding is driven by accent prestige (Giles, 1970, 1973). Naturally, lower prestige is associated with lower evaluations of the speaker across multiple dimensions. An IAT test conducted on Western Canadian participants showed that they implicitly preferred Italian to Mandarin Chinese accents in English (despite the lack of preference in explicit evaluations), which suggests that Italian accents invoke a more positive affect (and may carry a relatively high prestige) in this part of Canada. A study on the U.S. population by Lindemann (2005) also showed that the Italian accent was rated the most positively overall. Greek accents are likely to exhibit the same pattern. Thus, an empirical question to be addressed in future research is whether the effect found in Experiment 2 would replicate when a different non-native accent is used, particularly that of a generally stigmatized group (e.g., Mandarin Chinese or Indian, see Lippi-Green, 2012). It is important to note that despite this positive accent attitude found in previous research, both non-native speakers were still evaluated lower on status and solidarity traits than the native speaker. Importantly, group manipulation change seems to have been successful for truth ratings. It is thus unclear how successful group manipulation would be if another non-native accent was used, particularly one that is less preferred.

To date, the relationship between implicit and explicit attitudes remains a topic of debate. A meta-analysis based on 126 studies showed that the two tend to be positively correlated. Notably, this correlation strengthens when the two measures share the same underlying construct and when

people respond with a higher level of spontaneity in explicit surveys (Hofmann et al., 2005). Nevertheless, university undergraduate samples, in particular psychology students, may show a strong inclination to appear unbiased when responding to socially charged questions in a survey (i.e. regarding groups that face persistent discrimination) (Greenwald et al., 2009). Thus, the IAT can serve as a valuable psycholinguistic tool for examining how the processing of foreign-accented speech—or language that is merely attributed to non-native speakers in a written paradigm—is influenced by the strength of the implicit bias against foreign-accented speakers. Previous research has already shown, for instance, that more negative attitude toward other ethnicities is correlated with a poorer ability to correctly transcribe foreign-accented speech (which is not necessarily causal and one-directional since poorer ability to understand foreign-accented speech may foster more negative attitudes toward foreign-accented speakers) (Fraser & Kelly, 2012). Nonetheless, this avenue of research needs further exploration. A large effect size, the reliability of the effect, and the participant's inability to control their responses make the IAT a good candidate for examining individual differences in language processing, in line with mounting research showing meaningful differences between non-clinical adults (Hubert Lyall, 2019; Van Berkum et al., 2009b; van den Brink et al., 2012b).

Naturally, our study was subject to certain methodological constraints and limitations that future research could address. First, the simultaneous collection of EEG data necessitated a relatively modest convenience sample size. Notably, however, the sample exhibited considerable linguistic diversity, with nearly half of the participants being bilingual. Subsequent studies, particularly those not employing simultaneous EEG recording, may benefit from recruiting larger and more diverse participant pools, especially through the usage of online experimental platforms. Second, future studies could also examine a broader range of individual differences and demographic variables. For instance, we did not collect the second language of our bilingual participants, as this information could not have been meaningfully analyzed given our sample size constraints. However, future research may want to investigate this factor, since it has been suggested that proficiency in a minority language (for instance, an indigenous language in Canada) may enhance participants' sensitivity to group memberships and their societal implications (Ryan, 1983). The present study also did not collect racial demographic information from participants. Since group membership is closely linked to status effects, it is plausible that individuals from diverse racial

and ethnic backgrounds may respond differently to the stimuli and group membership information. This consideration is particularly salient given that all the speakers in the experiment were (purposefully) white women. In addition, future research may examine other native English accents, such as Australian or British, and compare them to non-native accents. Since both these groups of speakers would be immigrants to a particular region, it would be interesting to know how native accents associated with other countries affect domain-specific credibility.

5.6 Conclusions

Our findings contribute to a more nuanced understanding of the "incredibility effect" associated with foreign accents by examining a subtype of general world knowledge, namely, domain-specific "ingroup" knowledge. While we found some support for the original hypothesis that foreign accents may negatively impact perceived truthfulness, our results also reveal circumstances where foreign accents may enhance credibility, particularly when combined with signals of cultural integration. Crucially, neither study provided evidence for the fluency-based account, thus corroborating the outgroup-based framework. In addition, we found that there exists a reliable implicit bias against foreign-accented speakers within the knowledgeability domain, even in multicultural egalitarian societies such as Canada.

5.7 Ethics

All participants gave consent to participating in the experiment after they were provided with a full description of the study and were at liberty to withdraw from the experiment at any point for any reason with no consequences. The plan for this study was reviewed for its adherence to ethical guidelines by a Research Ethics Board at the University of Alberta (reference number Pro00123766).

5.8 Acknowledgements

This research was supported by a Social Sciences and Humanities Research Council of Canada Partnership Grant (Words in the World, 895-20161008). We extend our deepest gratitude to Taylor Lundstrom for her dedication and hard work in helping us collect the data for these studies. We

also thank Claire Weidemann, Gwen McGuire, Elizabeth Wood, Maureen Luckwell, and Victoria Blair for their kind assistance in data collection.

Chapter 6

General discussion

This dissertation investigated speaker and listener effects in the comprehension of language associated with interpretative uncertainty, either linguistic or epistemic in its nature. The first research goal was to investigate whether non-native speaker identity affects pragmatic interpretation of figurative language and facts of uncertain truth, i.e. speaker effects. We hypothesized that such statements by non-native speakers would systematically receive lower ratings from native speakers due to expectations regarding their LX proficiency, reduced shared knowledge between native and non-native speakers, more negative affect toward non-native speakers due to their outgroup social category, as well as reduced processing fluency due to phonetic deviations of non-native speech. The second research goal was to examine whether these judgments are affected by the listener's personality, specifically by traits that shape a person's response to the outgroup as well as general cognitive and psychological traits, i.e. listener effects.

While the effects of non-native speaker identity have been investigated before in psycholinguistics, the overwhelming majority of studies have focused on the acoustic processing of non-native speech (Adank et al., 2015; Clarke & Garrett, 2004; Van Engen & Peelle, 2014, among many others) or the processing of syntactic errors and semantic non-sensicalities in different accents (e.g., Caffarra & Martin, 2019; Grey & van Hell, 2017; Hanulíková et al., 2012; Holt et al., 2018). A slightly smaller body of research has investigated prediction during the processing of non-native semantically well-formed sentences (e.g., Goslin et al., 2012; Porretta et al., 2020; Romero-Rivas et al., 2016). Very few studies have investigated figurative language from this perspective (Bazzi et al., 2022; Caffarra et al., 2018, 2019). The domain of unknown facts has received more attention (Foucart et al., 2019, 2020; Foucart & Hartsuiker, 2021; Lev-Ari & Keysar, 2010; Lorenzoni et al., 2022; Meo et al., 2011; Podlipský et al., 2016; Souza & Markman, 2013; Stocker, 2017; Wetzel, 2021) but has been plagued with inconsistent findings.

The aim of the present dissertation was to contribute to the previous research by investigating non-native speaker effects in the perception of irony, metaphors, and domain-specific facts of uncertain truth. The studies in this dissertation were carefully designed to compare different types of irony

and metaphors with literal sentences, whereas generally unknown facts were compared with facts generally known to be true or false. In addition, this dissertation contributes to a very limited body of literature on listener effects in non-native speech processing, as well as offers the first investigation of generalized implicit accent prejudice as a predictor of linguistic judgments during foreign-accented speech processing.

6.1 Summary of the results

A brief overview of the results is provided in Table 6.1.

Table 6.1. A brief summary of the results of the dissertation.

Experiment #	Modality	Lower ratings of non-native speech	Affected aspect	Individual differences modulating the results
Experiment 1 (irony)	auditory	Yes	Ironic criticism and praise (not literal language)	Political conservatism
Experiment 2 (metaphors)	written	Yes	Both metaphorical and literal sentences	Bilingualism Linguistic diversity Cognitive reflection Extraversion Political conservatism/ Explicit accent attitudes
Experiment 3 (world knowledge)	written	Maybe	Unknown facts (not true or false facts)	-
Experiment 4 (world knowledge)	auditory	No	True, false, and unknown facts	Bilingualism Implicit accent attitudes

6.2 Speaker effects

6.2.1 Linguistic uncertainty

Listeners have been shown to incorporate various aspects of the speaker's identity during language comprehension (e.g., Bergen & Grodner, 2012; Bornkessel-Schlesewsky et al., 2013; Gibbs et al., 1991; Jiang et al., 2020; Kroczek & Gunter, 2021; Regel et al., 2010, 2010; Van Berkum, van den Brink, et al., 2008). Figurative language, due to its open-endedness and ambiguity, may be particularly susceptible to speaker effects (e.g., Gibbs & Colston, 2012; Regel et al., 2010). As discussed extensively in the Introduction, since non-nativeness is often associated with less

accuracy in semantic choices due to not having full linguistic and communicative competence as well as with reduced shared knowledge, listeners might rely on these assumptions about the speaker when they are dealing with pragmatic ambiguities—namely, when encountering statements that in their surface form are counterfactual to the preceding context (like ironies) or seemingly anomalous (like novel metaphors). Specifically, listeners may find it more difficult to derive a figurative meaning, feel less confident in their interpretation, perceive the meaning as making less sense or being less appropriate when they know that the speaker is non-native. In short, the findings of the first two experiments support this hypothesis.

In Experiment 1, we built upon two previous studies (Caffarra et al., 2018, 2019) to examine whether naturalistic prosody and peer relationship between interlocutors would facilitate meta-representational inferences from non-native irony for native listeners. Crucially, we used multiple rating scales (irony, certainty in interpretation, appropriateness, and offensiveness) to get a multi-faceted insight into the phenomenon and explore the potential consequences that may arise if irony is misunderstood. Notwithstanding the peer status of the interlocutors and in partial contrast to the findings of Caffarra et al. (2018), we found that *both* types of irony, the more common criticism and the less common praise, were rated as less ironic when delivered in a non-native compared to a native accent. Both types of non-native irony also evoked more doubts about the speaker's intent than native irony, although the effect was marginal for ironic praise. This may suggest that no matter who produces it, ironic praise evokes a certain level of uncertainty about the speaker's intentions, perhaps due to a surprisal effect from it being rarer and more unnatural. Therefore, the effect of accent here was less salient. I will return to this finding when discussing Experiment 2 as well. These results are unlikely to stem from intelligibility issues per se because the results of the intelligibility pre-test did not significantly correlate with either irony or certainty ratings, and because literal comments in both accents were rated equally low for irony. They, however, may still stem from lower comprehensibility[19], since there is no easy way to gauge it. There was no effect of non-native accent in ratings for appropriateness and offensiveness of ironic remarks, although non-native literal criticism was rated less offensive than native literal criticism. This may

[19] I follow definitions by Munro & Derwing (1995) where comprehensibility refers to the subjective *ease* of understanding a non-native utterance whereas intelligibility denotes how well the utterance was actually understood objectively (most often assessed by asking people to provide an orthographic transcription of the utterance).

be because listeners assume different cultural norms for using explicitly critical remarks and might conclude that the speaker did not intend to be rude, thus giving them the benefit of the doubt (Gibson et al., 2017). As predicted in the dissertation Introduction, accent had no effect on irony ratings for literal conditions (cf. Bazzi, Brouwer, & Foucart, 2022 who found both ironic and literal praise to be rated lower for irony when attributed to a non-native speaker). Note that since irony seems to be a universal phenomenon and since it does not require any *linguistic* skills over and above what is required for literal language barring prosodic marking (e.g., "You are a really *good/bad* cook!"), we consider the explanation to be improbable that a non-native speaker who uses complex, grammatically correct sentences (in all four studies in the literature as of today) would not have enough linguistic competence to make use of irony. What is more likely is that non-native speakers may generally avoid using irony in conversations with native speakers due to multiple factors, such as risk of misinterpretation due to challenges with conveying it prosodically, fear of using it in culturally inappropriate circumstances, or not feeling close enough to their interlocutor as they tend to *expect* stigmatization from others on the grounds of their accent (see Derwing, 2003; Gluszek & Dovidio, 2010). Linguistically speaking, even under the standard pragmatic model positing that literal meaning needs to be rejected first, it is arguable that the literal meaning was—partially or fully—rejected here, as non-native irony was still rated much higher for irony than respective literal sentences (5.6 vs. 2.9 for criticism and 4.8 vs. 2.2 for praise). This raises a point that treating linguistic judgments as binary (acceptable/unacceptable or ironic/literal) risks losing valuable information, since fine-grained yet statistically significant distinctions may be lost (see section 6.4.2 for more discussion on the subject). All in all, it appears that despite being able to tell apart literal and ironic utterances in both accents, native listeners succumb to the illusion that non-native ironies are inferior to native ironies—likely because they merely *expect* them to be due to factors extensively discussed above.

In Experiment 1, we only analysed a few global acoustic measures of the recorded stimuli calculated over entire utterances and found no interactions between irony and accent. To examine whether this interaction was present in other measures and at different levels of granularity (vowels, syllables, words, multi-word intervals, and whole utterances), we conducted a fine-grained acoustic analysis reported in Chapter 3. In addition, we wanted to investigate whether ironic praise and ironic insults in general were associated with distinct prosodic patterns. Our

hypothesis was that Mandarin-accented irony would be more prosodically contrasted with literal speech than native irony (indicating compensatory behaviours), and that ironic praise would be more prosodically marked compared to ironic criticism. The results showed that Mandarin-accented irony was indeed prosodically marked to a larger extent than native irony with respect to almost all pitch-related measures. In addition, we found significant differences between native and non-native irony in vowel duration. Whereas native irony of both subtypes had a longer mean vowel duration than literal speech (see Anolli et al., 2000; Haiman, 1998), Mandarin-accented ironic praise had a much longer mean vowel duration than ironic criticism (which did not differ from literal conditions). This confirmed our prediction that non-native speakers may put extra effort into marking ironic praise prosodically because the social cost of misinterpreted ironic praise is too high. The fact that ironic criticism did not differ from literal conditions may be explained by the fact that it is perceived as less threatening and critical than literal criticism (Brown & Levinson, 1987), hence potential misunderstanding here carries no social cost. While a linguistic transfer cannot be completely ruled out due to the lack of studies on different types of ironies in native Mandarin that compare them to the same baselines that we used, several factors lend support to the exaggeration hypothesis (for instance, much higher pitch variability in the target interval in non-native than in native speech, whereas a linguistic transfer might be expected to be more consistent). The findings of this analysis do not necessarily indicate that the results of Experiment 1 stem purely from these acoustic markers—even more so considering that non-native ironic praise was rated as less ironic even when flat intonation was used (Caffarra et al., 2018) or no speech at all (Bazzi et al., 2022). What they do suggest, however, is that non-native speakers may be indeed less comfortable using ironic praise than native speakers—or, possibly, irony in general. This would be in agreement with previous literature suggesting that when cultural knowledge is not shared, listeners experience significantly more trouble with identifying ironic intent (Kreuz & Link, 2002), and speakers tend to use less irony overall (Averbeck, 2015; Averbeck & Hample, 2008). Based on this, listeners may not *expect* non-native speakers to use it as often as native speakers do.

Experiment 2 aimed to investigate whether metaphorical sentences by native and non-native speakers are perceived differently. To the best of our knowledge, no previous research has tackled this question. We investigated two types of metaphors, familiar and unfamiliar/novel ones,

comparing them with closely matched literal counterparts. We did not ask the participants to rate the sentences for their metaphoricity to avoid varying perceptions of what constitutes a metaphor and to prevent drawing attention to the fact that we are investigating metaphors at all (note that sensibility ratings in metaphor research are common, see, for instance, Al-Azary & Buchanan, 2017; Gibbs et al., 1991). What we collected instead were judgments of how much sense the sentence makes as a proxy of comprehensibility and perceived semantic coherence. Since our acoustic analysis revealed hidden segmental and prosodic confounds in non-native speech, we opted for an exclusively written design. The results suggested that all sentences attributed to a non-native speaker were perceived as less sensical, mirroring the "illusion of ungrammaticality" reported for grammatical errors (Wesołek et al., 2023) but in the pragmatic domain, and took significantly longer to evaluate. The reason for this effect is likely to be expectations of lower LX command, which may be additionally mediated by lower accent prestige and thus lower social status (although we did introduce both native and non-native speakers as undergraduate students from the same university, which should have reduced to some extent the influence of such stereotypes). Since no actual speech was heard, the results are unlikely to stem from extra cognitive load (but see 6.4.2 for a discussion) and rather have to do with expectations induced by describing a speaker as strongly accented. While there was only a main effect and no significant interactions, some of it is likely due to ceiling effects observed mostly in the literal and to some extent in the familiar metaphor condition that reduced variability by compressing responses. Both unfamiliar metaphors and ironic praise in Experiment 1 also produced the largest standard deviations of all conditions in the two experiments. This shows that rarer linguistic tropes exhibit less clear speaker effects as their lower naturalness and frequency shift the focus away from the speaker's identity and toward the novelty of the language itself. For psycholinguistic research on the effects of non-native speaker identity, this counterintuitively means that extreme linguistic constructions do not always produce the strongest results, and more common constructions that are not borderline semantic anomalies may be more informative, in no small part due to interpersonal variation in their processing being less profound. Overall, while no interactions between speaker and metaphor conditions were found, which contradicts our initial hypothesis that literal sentences may remain unaffected by non-nativeness, such results are not unheard of. Both literal and metaphoric sentences were rated as less meaningful when the participants believed they were created by a computer with a random generator as compared to a poet in Gibbs et al. (1991), and both literal

and ironic sentences were rated lower for irony when attributed to a non-native than native speaker in Bazzi et al. (2022). Hence, speaker-induced expectations may profoundly affect subsequent language processing regardless of the linguistic phenomenon, with listeners/readers anticipating less sensical semantic constructions from a random generator or from a non-native speaker, which makes them process their statements with more skepticism. The fact that mean rating time was also longer for the non-native speaker suggests that assumed non-nativeness increases cognitive load and makes pragmatic inferencing more taxing even with no oral speech.

All in all, the two experiments convincingly show that native listeners/readers do take the identity of the speaker into account when making linguistic judgements about figurative language. At least for judgments examined here (sensibility, degree of irony, certainty in speaker's interpretation), this entails negative consequences for non-native speakers, namely that their speech is perceived as less sensical and less ironic. These consequences arise not only due to disrupted processing fluency but also due to mere *expectation* of lower language proficiency, as demonstrated by using auditory and written modalities. Since both ironic and metaphorical language are extremely common and constitute a large share of daily conversations (Gibbs, 2000; Hancock, 2004; Steen et al., 2010; Whalen et al., 2009), the findings of this dissertation may indicate tangible consequences for billions of LX speakers worldwide.

6.2.2 Epistemic uncertainty

As discussed in the Introduction, non-native accents disrupt processing fluency, and processing fluency has been shown to affect resolution of epistemic uncertainties. As anticipated based on these findings, Lev-Ari & Keysar (2010) reported that generally unknown facts (e.g., "Ants don't sleep") are believed less when they come from non-native speakers. They also explicitly attributed this finding to processing fluency only since the non-native speakers in their experiment were introduced as simply reciting facts the experimenter gave them, i.e. the *fluency-based hypothesis*. However, numerous studies failed to replicate this finding in full (Foucart et al., 2019, 2020; Foucart & Hartsuiker, 2021; Lorenzoni et al., 2022; Meo et al., 2011; Podlipský et al., 2016; Souza & Markman, 2013; Stocker, 2017; Wetzel, 2021). Some studies did not replicate it at all even using the exact same design (Souza & Markman, 2013), other studies showed differences only in neural measures, even without any oral speech (Foucart et al., 2019, 2020; Foucart & Hartsuiker, 2021),

and some even showed that non-native speakers may be considered *more* credible (Lorenzoni et al., 2022). The original study was recently replicated and extended by the same author (Boduch-Grabka & Lev-Ari, 2021), but the observed results did not reproduce in a close preregistered replication with the same auditory materials in another lab (Barlow et al., 2024). Crucially, Barlow and colleagues also did not reproduce the finding that exposure to Polish-accented English improves veracity judgments of unknown Polish-accented statements. Although the authors argue that this does not imply that the processing fluency account is incorrect (since both accents in their study were rated as equally easy to understand), these null results cannot be taken as direct support for this account either. This wide variety of effects is in no small part due to methodological and conceptual inconsistencies between the studies. For instance, to the best of our knowledge, only Foucart and colleagues pre-tested their stimuli to ensure that statements were correctly assigned into their respective true, false, and unknown conditions (Foucart et al., 2019, 2020; Foucart & Hartsuiker, 2021). A few studies used a post-test (by asking the participants after the experiment whether they knew a particular fact) (Lev-Ari & Keysar, 2010; Stocker, 2017). Several studies took most of the stimuli from Lev-Ari & Keysar (2010) but supplemented them with their own stimuli, without a pre- or post-test (Hanzlíková & Skarnitzl, 2017; Podlipský et al., 2016; Souza & Markman, 2013) and some seem to not have checked the participants' familiarity with the facts at all (e.g., Lorenzoni et al., 2022). The vast majority of the studies did not use true and false controls. In addition, some studies were in-person and some were online—for those that were online (Barlow et al., 2024; Boduch-Grabka & Lev-Ari, 2021; Lorenzoni et al., 2022; Meo et al., 2011; Stocker, 2017), it is not clear whether participants could just google the correct answer or not. Populations and accents were vastly different as well. Hence, the lack of converging findings is actually not surprising, and the so-called "incredibility effect", if exists at all, is clearly not very robust. Moreover, the fact that differences between native and non-native speakers were observed without any oral speech coupled with the fact that raising the social status of the speaker improves credibility (Foucart et al., 2019) strongly suggests that social categorization plays a substantial role in the potential "incredibility effect", i.e. *the outgroup-based hypothesis*.

This dissertation aimed to contribute to previous research by examining domain-specific instead of domain-general knowledge using a different form of social category and status manipulation—presenting non-native speakers as either Canadian citizens or Immigrants. Since the original

findings on domain-general distrust toward non-native speakers have been hard or impossible to replicate, we decided to employ a stronger manipulation of using domain-specific stimuli. This manipulation is stronger because listeners might naturally assume that foreigners know less about certain "ingroup" topics and more about certain "outgroup" topics (Fairchild & Papafragou, 2018, p. 90), although this has not been demonstrated empirically before. We focused on obscure facts about a particular place (i.e. Canada) that participants were not expected to know, such as the tenth largest city, the number of national parks or the exact population size in a particular province, history facts, geographical locations of small towns, etc. To the general population, this "Wikipedia-depth" knowledge is of very low relevance, and most people would not come across or remember it unless they have a specific interest. We also hypothesized that making non-native speakers a partial ingroup by labelling them as Canadian citizens would raise their credibility on the ingroup knowledge compared to labelling them as Immigrants, both because it may raise their social status and because it is usually associated with longer duration of living in the country (the exact duration of living, however, was not specified). Additionally, we examined how much of the effect stems from the auditory signal itself by comparing written (Experiment 3) to auditory (Experiment 4) presentation of materials. Overall, we found no evidence for the fluency-based hypothesis.

Experiment 3 found very marginal evidence that non-native speaker identity is taken into account in the written modality. Unknown statements attributed to both foreign-accented speakers (regardless of their group membership) were perceived as less true than the same statements attributed to the native-accented speaker. This effect emerged only for unknown sentences and was significant in the main model but not in the post hoc comparisons. Considering inconsistent findings observed in previous research, numerous failed replication attempts, and the fact that the outcome of this line of research has profound societal implications, it is vital that the results of this study are replicated on a much larger participant sample. The sample limitation in these studies was due to the simultaneous EEG recording.

In Experiment 4, however, a completely different pattern emerged. *All* statements (true, false, and unknown) were rated higher for truth when the speaker was foreign-accented and introduced as Canadian. We explain this "knowledgeable immigrant" effect by a mixture of the modality and the nature of our stimuli. Since all the experimental sentences were about Canada and since a

citizenship test in Canada requires knowledge of such facts, it might have made the Canadian foreign-accented speaker the most credible among the three speakers. In a nutshell, the accent serves as an auditory reminder of the speaker's recent effort to integrate into Canadian society. The fact that this effect only emerges under auditory stimuli suggests that explicit auditory cues are needed for the group manipulation to be successful—i.e. group membership change is modality-specific.

The modality-specific results themselves are actually not surprising. Prior literature has shown that people process information in the written and auditory modality differently. For instance, auditory presentation leads to better memory retrieval (Markman et al., 2007) and better memory for words in general (Conway & Gathercole, 1987), affects the resolution of temporal but not semantic anaphora (Jakimik & Glenberg, 1990), and helps to understand complex texts better (Carroll & Korukina, 1999). It is also reasonable to assume that social characteristics of the speaker are also more salient in oral speech. Perhaps this is the reason why group membership change was more successful. Attentional effects may also be at play. With written sentences only, participants may be focusing largely on the most salient feature (presumably the accent heard *prior* to the experiment) while disregarding the visual cues (flags) indicating group membership. This idea has empirical support from earlier studies, as, for instance, Foucart et al. (2019) reported an early negativity associated with just looking at the face of non-native as compared to native speakers in the written modality. All in all, these findings strongly indicate that modality can be of major importance for the processing of psycholinguistic stimuli, especially when non-native accents are involved. More research on non-native accents needs to employ the comparison of the two types of modalities, specifically to dissociate the effect of reduced processing fluency from the effects of social categorization and stereotyping.

All in all, the two experiments do not provide robust evidence that non-native speakers are trusted less on domain-specific facts of unknown truth value. The findings from the written modality are inconclusive due to marginal significance in post hoc comparisons, and the findings from the auditory modality indicate an *advantage* of being a foreign-accented speaker when it is combined with a clear indication of their efforts to integrate into the country. Importantly, true and false statements were equally affected in the latter case, which contradicts our original hypothesis.

6.2.3 Social evaluations of the speaker

Social evaluations of the speaker were only collected after Experiment 2 (metaphors in the written modality) and Experiment 4 (world knowledge in the auditory modality). Not collecting them after Experiment 3 was solely my omission. In Experiment 2, social evaluations of native and non-native speakers did not differ for either the overall affective evaluation ("How do you feel about this person overall?") or the solidarity dimension of likeability. These results are consistent with Bazzi, Brouwer, & Foucart (2022) who found no differences in speaker ratings for affect, status, or solidarity using a written design (cf. Lorenzoni et al., 2022). In Experiment 4, however, the native speaker was consistently rated higher than both non-native speakers on every dimension: the overall affective evaluation, the status dimension of knowledgeability, and the solidarity dimension of pleasantness. Since we did not collect speaker evaluations after Experiment 3, which was identical to Experiment 4 except for modality, we cannot directly link stimulus modality to social evaluations of the speaker. However, the results of Experiment 2 and 4 together, as well as findings by Bazzi, Brouwer, & Foucart (2022) do underscore the impact of perceptually salient auditory cues on social evaluations. Auditory characteristics of the stimuli likely activate and/or reinforce social stereotypes related to speakers with non-native accents more readily than abstract verbal labels and written vignettes. Another possibility is the degree of conscious control used during the rating phase and proneness to self-presentation strategies that may be different between the two modalities. Specifically, without the direct auditory cue of a non-native accent, participants may engage in a more controlled, reflective mode of responding and become more aware of the potential social biases at play, adjusting their responses accordingly.

The crucial question, however, is whether this lack of difference in explicit social evaluations in the written modality suggests that social stereotypes are less, or not at all, activated and thus do not affect linguistic processing in that modality. This is a difficult question to answer. On the one hand, we did find non-native sentences to appear less sensical than native sentences, whereas Bazzi, Brouwer, & Foucart (2022) found all non-native sentences to appear less ironic than native sentences. On the other hand, our results do not need to stem from social stereotyping across status or solidarity dimensions—they may simply stem from the expectation of lower LX proficiency. Of course, expectation of lower LX proficiency is also a form of stereotyping since accent strength and proficiency should not be confused (Gluszek & Dovidio, 2010). More research is thus needed

to shed light on this issue. A more general discussion about the use of written vs. auditory modality is available in section 6.4.

6.3 Listeners effects

Research in psychology has convincingly demonstrated that individual differences do not have a uniform effect across different situational contexts (Mischel, 2013). Meanwhile, some research in psycholinguistics has delivered converging findings that individual differences and situational contexts dynamically interact, and that not all linguistic variables are affected by everything (see, for instance, Gucman, 2016 who examined ratings of ironic and literal sentences on multiple scales and how they interact with individual and social factors). Likewise, this dissertation did not find a single character trait that affected pragmatic processes across all the studies. Rather, different traits were more important for different types of language examined, either interacting with non-nativeness or being a thing on their own. While extensive discussions of individual differences are available after each study reported in this dissertation, the section addresses the overarching theme and what the results mean for the big picture.

In Experiment 1, only political orientation was predictive of irony comprehension, so that higher conservatism was associated with lower comprehension accuracy. This effect, however, did not interact with non-nativeness but rather was consistent across both speakers. Both types of ironic comments were rated as less ironic and literal praise was rated as more ironic by more conservative compared to more liberal participants. Affective and cognitive empathy were not predictive of irony comprehension, and neither was discomfort with ambiguity (need for cognitive closure). This suggests that it was not the varying levels of discomfort with ambiguity associated with political views that affected irony comprehension, since a targeted measure of need for cognitive closure was not predictive. This finding emphasizes that when examining the effect of political views in particular, it might be important to simultaneously examine traits that have been found associated with political orientation, such as openness to experience (Sibley et al., 2012), need for cognitive closure (Chirumbolo et al., 2004) and other relevant traits, to make sure that political views affect the results over and above these correlated variables. Of course, it is impossible to ensure that any traits found predictive (by "predictive" I, of course, mean "associated", because causality is impossible to establish in correlational studies) are actually predictive by themselves

and not through other mediating variables. Careful reasoning based on theoretical frameworks and examination of other traits relevant to the linguistic phenomenon under consideration should help to reduce the effect of such confounds. That said, the results of this study might be driven not by other mediating variables but rather by ontological differences, i.e. different premises about what constitutes an ironic statement between left- and right-leaning participants. This is an important empirical question that is hard to investigate and likely requires extensive qualitative studies. Some of the effect of political orientation on irony comprehension might be reduced by using statements created by people who fall in different places on the political spectrum, which was not the case in this experiment. Follow-up research is thus needed to get to the core of this effect.

Experiment 2 delivered the most robust correlations between metaphor comprehension and individual differences. In part, this might be explained by a very diverse participant sample that ensured an excellent score distribution for many measures that were of interest to us. Such variables included the linguistic diversity of a participant's environment, extraversion, cognitive reflection, and political views. Apart from the above-mentioned fact that the sample for this experiment was really diverse, which allowed for better examination of individual differences, it may also be due to the non-specific nature of the task (sensibility judgments) and the phenomenon under consideration (metaphors). All in all, the results of this study indicate that there is substantial individual variation in linguistic judgments that needs to be accounted for. Most traits that were found to be significant interacted with the most unnatural type of metaphors (novel ones), which is to be expected as rarer linguistic tropes are often subject to the largest interpersonal variation. Future studies may examine the time-course of these effects.

Experiments 3 and 4 were somewhat unlucky with respect to individual differences for several reasons. First, as EEG data was simultaneously collected, this severely limited the sample size and its diversity. Small samples are naturally less informative with respect to individual differences, but convenience samples are even less so. Second, these two experiments were conducted in-person as compared to the previous two studies that were done online. As discussed extensively in the Introduction (see also section 6.4 for a detailed discussion of methodology), participants in in-person experiments may be more strongly subject to the social desirability bias, providing answers in surveys that are different from their honest opinions. This would explain extremely poor internal reliability of several surveys in Experiments 3 and 4 as measured by Cronbach's α (see Table 6.2)

or extreme skewness of their scores toward the "socially acceptable" answers (specifically, Wilson-Patterson political ideology, the Social Dominance Orientation, and the explicit Accent Attitude Scale). As but one example, Cronbach's α was 0.65 for the Wilson-Patterson Conservatism scale in in-person Experiment 3 but 0.79 in online Experiment 1, although both studies recruited university students. One can argue that sample sizes differed between the two experiments. Although the sample sizes were indeed different, the effect of the sample size is usually modest compared to other factors, such as the number of items in the scale and their intercorrelations (see Bonett, 2002).

Table 6.2. Internal reliability (Cronbach's α) of the individual difference measures used in each study.

Survey	**Exp 1 (online) N = 93**	**Exp 2 (online) N = 85**	**Exp 3 (in-person) N = 30**	**Exp 4 (in-person) N = 33**
Wilson-Patterson conservatism	0.79	0.90	0.65	0.72
Need for cognitive closure	0.83	N/A	N/A	N/A
Interpersonal reactivity index	0.85	N/A	N/A	N/A
Accent attitudes scale	N/A	0.76	0.51	0.64
Big-V extraversion	N/A	0.90	N/A	N/A
Big-V agreeableness	N/A	0.69	N/A	N/A
Big-V conscientiousness	N/A	0.83	N/A	N/A
Big-V neuroticism	N/A	0.93	N/A	N/A
Big-V openness	N/A	0.83	N/A	N/A
Social dominance orientation	N/A	N/A	0.89	0.90

The Implicit Association Test where the participants could not consciously control their responses was one of the only two measures we could use in Experiments 3 and 4. In addition, in line with predictions by Ryan (1983), the participant's bilingualism significantly modulated the results, so that only monolingual speakers exhibited the main effect found in the study. All in all, the data we collected about how individual differences interact with the resolution of epistemic uncertainty is very limited. Future research that does not employ simultaneous EEG collection may focus on larger and more diverse participant samples to investigate this issue.

6.3.1 Implicit Association Test as a measure of implicit accent attitudes

Experiments 2-4 delivered converging findings that non-native speakers are implicitly considered less knowledgeable than native speakers. These findings agree with Roessel et al. (2018) in both

magnitude and directionality. Methodologically, this means that the overall effect is replicable in different societies. All three tests in this dissertation produced an effect of a large magnitude. However, some general limitations and criticisms of the test—and the concept of implicit attitudes itself—need to be discussed.

To begin with, prior research suggested that the IAT responds to things like the setting in which it is taken, the specific instructions given to the participant, unrelated tasks that precede the test, or a person's recent experience of intergroup contact (e.g., Dasgupta & Greenwald, 2001; Han et al., 2010; Lowery et al., 2001). For this dissertation, this could mean that the test scores were affected by the experiment proper that preceded it. We consider it unlikely for several reasons. First, the three experiments produced results of similar magnitude, even though the tasks differed (metaphor vs. world knowledge processing and evaluation) and modalities differed (written vs. auditory). Numerically, the largest effect size (Cohen's d = 1.36) was obtained in the written modality in Experiment 2 on the U.S. population. Second, when the results of the IAT were affected by prior tasks in previous research, those prior tasks usually involved affective content, such as faces of admired/disliked individuals of different races (Dasgupta & Greenwald, 2001) or stories where a particular group behaved negatively or positively (Han et al., 2010). Since we deliberately tried to minimize differences between native and non-native speakers in all respects but their accents, counterbalanced faces/biographies, and did not portray any speaker in a negative light, any effect of the experimental task on the IAT scores is unlikely. As far as the order of tasks goes, arguably, administering the IAT before the experiment would have had a more detrimental effect, since subsequent linguistic processing might have been affected by (re)-activating the stereotypical associations with native and non-native speakers (i.e. conceptual priming).

In addition, the IAT has been criticized for its low test-retest reliability (inconsistent results when the test is taken multiple times by the same person). For instance, when the test is administered twice on the same person, some studies show test-retest reliability at weak to moderate levels (K. A. Lane et al., 2007). The low reliability may raise questions about whether the test can accurately assess stable, long-term implicit biases in an individual. Some researchers argue that imperfect test-retest reliability is more about the validity of the scoring procedure than it is about the validity of the method itself (Richetin et al., 2015). That said, D-scores computed in this dissertation are still considered the second-best scoring procedure, slightly inferior only to a more complicated

alternative (DW-scores; see Richetin et al., 2015), which improves the test-retest reliability but not dramatically (Chevance et al., 2017).

And finally, what the test measures conceptually and how implicit attitudes are related to explicit ones remains debated. For instance, some argue that it simply measures cultural awareness of existing stereotypes via the workings of the predictive brain rather than personal bias (Hinton, 2017; Olson & Fazio, 2004)—i.e. a person may be aware of stereotypes without personally endorsing them. Some researchers criticize this idea and suggest that personal and cultural associations are represented similarly in memory, with personal relevance emerging from conscious endorsement (Gawronski & Bodenhausen, 2006). Nevertheless, we did find the test to be predictive of linguistic judgments, even if the individual may not endorse such stereotypes. And, of course, debates are ongoing as to whether implicit and explicit attitudes are correlated, and which one is a better predictor of behaviour. A meta-analysis showed that the two tend to be positively correlated (Hofmann et al., 2005). Greenwald & Banaji (2017) argue that weak correlations still tend to be fairly common and offer three possible explanations for this: 1) implicit and explicit attitudes tap into different constructs; 2) they are susceptible to situational effects differently; and 3) at least one of them, which is likely to be self-report, lacks validity. A meta-analysis the authors carried out suggests that (3) is especially pronounced for socially sensitive topics where self-report is unreliable, and the predictive validity of implicit attitudes in that case is far superior (Greenwald et al., 2009). Precisely this is the reason why this dissertation examined both explicit and implicit attitudes, and it may be the reason why only the implicit attitudes were predictive in Experiment 4. Importantly, whereas Experiments 3 and 4 only differed in their modality, implicit attitudes were only predictive when the stimuli were aural. This may suggest a possible limitation of using the IAT in the written paradigm. Whereas the IAT scores were predictive in Experiment 2 that also used a purely written modality, the scores did not interact with the speaker condition, which reinforces the point made above. In addition, explicit attitudes were also predictive in Experiment 2. As it was conducted fully online, it was likely less susceptible to the social desirability bias compared to in-person experiments when the experimenter and research assistants were present in the other room (and likely believed to have access to your responses). Of course, this assessment remains speculative as participant samples in these experiments cannot be compared directly, with the sample in Experiment 2 being much more diverse in all respects

than the undergraduate samples in Experiments 3 and 4. Future research should directly investigate this question by comparing online to in-person administration of the same experiment using the undergraduate university pool. This brings us to a more general discussion of the problem of sampling and methodology in psycholinguistic research in general and this dissertation in particular.

6.4 Limitations and future directions

As with any empirical investigation, we acknowledge some inherent theoretical and methodological constraints of the approach taken in this dissertation. This section covers these shortcomings and suggests ways in which future studies may improve on them. It also discusses some general directions future research can take to investigate the effects of non-native speaker identity in communication.

6.4.1 Sampling and the testing environment

Unlike in psychology, very little research in psycholinguistics has been done on the effect of the testing environment on the results of the studies. Testing environment includes both the location where an experiment takes place (a university lab, a classroom, online on a personal computer) and the social setting in which it occurs (such as the presence or absence of other participants, social characteristics of an experimenter, etc.). Researchers often choose a particular method based on convenience and time/money constrains. Mass testing participants in the lab is quicker and easier than testing them individually, and online data collection is faster and requires fewer lab resources than the in-person one (Skitka & Sargis, 2006). The COVID pandemic imposed additional restrictions, such as a temporary prohibition on in-person testing (in fact, this was why Experiment 1 was conducted fully online). Although oftentimes it is assumed that the results between online and offline studies will be equivalent, it is, in fact, reasonable to expect that the testing environment can make a profound difference. First of all, when participants come to the lab for in-person testing, they may be subject to the social desirability bias discussed in detail in the Introduction. Although we cannot make any definite conclusions, but this seemed to have had a direct impact on the internal reliability of several individual difference questionnaires administered after the experiment. Specifically, as we discussed in section 6.3, internal reliability

plummeted in in-person experiments, likely due to the fact that participants trended toward "socially acceptable answers" rather than honest answers. Second, participants may believe that there are right or wrong answers to the experimental task, even if told otherwise, and modify their behavior accordingly. Barenboym et al. (2010) argued that rating tasks in psycholinguistics might be particularly prone to the environment effect and showed that ratings of words for danger and familiarity in online and in-person surveys differed significantly for certain stimuli. They also showed a potential gender effect in that women's in-person and online ratings correlated significantly better than men's. It is thus important to address the sampling and the testing environment of the studies in this dissertation.

Participant samples in all the reported experiments differed quite substantially. Participants for Experiment 1 were recruited online exclusively through convenience sampling from the undergraduate LING 101 pool at the University of Alberta and received partial course credit. Participants for Experiments 3 and 4 were all students at the University of Alberta who completed the studies in-person. Fifty-seven percent of participants in Experiment 3 and ninety-seven participants in Experiment 4 received a monetary reward; the rest received partial course credit. Finally, participants in Experiment 2 were recruited online through Prolific from the United States according to gender and political quotas. Also notably, platforms like Prolific have their own confounds since their participants essentially self-select and thus differ in many character traits (Douglas et al., 2023). This alone may explain a substantial amount of variation, particularly in the effects of individual differences. For instance, participants receiving a monetary reward might be more attentive and enthusiastic than those who participate for partial course credit (cf. Luccasen & Thomas, 2014). This may have contributed to the discrepancy between Experiments 3 and 4, although the modality of the stimuli likely contributed more. The monetary reward versus course credit issue also pertains to Experiments 1 and 2. In addition, as explained above, online and offline studies are also likely to be significantly different for many reasons, hence there is no guarantee that the results of the studies reported here would replicate in a different medium.

Some of it is reassuring, since online Experiment 1 and in-person studies done by other researchers (Bazzi et al., 2022; Caffarra et al., 2018, 2019) produced largely converging results, despite differences in physical testing environments, modality, and non-native accents used. Hopefully, Experiment 2 will be replicated in a different environment, as well as using a different modality

and different accents. The replicability of Experiments 3 and 4 is hard to assess, since previous research on world knowledge produced extremely conflicting results. If both processing fluency and the strength of social categorization contribute to the credibility of non-native vs. native speakers, the results may vary substantially with population, prestige of the accent used, speaker confidence and sex, the domain of world knowledge, and stimuli modality, among other things. We are not aware of any studies that directly compared written to oral delivery of non-native versus native stimuli, but the results of Experiment 3 and 4 reported here provide initial evidence that drastically different patterns may be observed in different modalities (see also Souza, 2012 for the effect of modality on processing fluency). Hence, ideally, both modalities should be simultaneously tested. Next section discusses other important considerations when evaluating the results of tasks in different modalities.

6.4.2 Modality

The written paradigm pioneered by Fairchild & Papafragou (2018) is becoming increasingly popular to separate processing confounds from other effects associated with non-nativeness. However, it is currently not clear whether that is really the case. Prior research has shown that readers do generate implicit prosody (inner speech) when reading, which affects the processing of written sentences (see Breen, 2014 for a review). People also read at a lower speed when asked to imagine the voice of a speaker who talks with a slower speech rate—although they do not seem to be affected by the linguistic identity of the speaker manifested in their native or non-native accent (Zhou & Christianson, 2016b, 2016a). Although the authors argued that such auditory simulations generate a more detailed prosodic representation of the text than the default prosodic contour, readers might still unconsciously generate "non-native" implicit prosody even without the cued simulation, in particular if the accent exposure pre-experiment is sufficient (or if the accent is played before every trial). This perceived non-fluency could affect the way readers process the sentence cognitively. To shed more light on this issue, future research could examine whether individual differences in inner speech vividness moderate the effects of non-native accents in written paradigms or whether explicitly describing the speaking style of non-native speakers affects written sentence processing.

6.4.3 Other methodological considerations

As is now clear, this dissertation focussed on offline linguistic judgments. As with any method of assessing the computed linguistic meaning and the computational process itself, linguistic judgments are not infallible. Participants have to reflect on their comprehension after reading the sentence, which does not capture any real-time processes. Ratings may be prone to the social desirability bias, particularly when native versus non-native speakers are involved. Ideally, a multi-methodological approach should be employed (Arppe & Järvikivi, 2007), including qualitative methods, to get a more comprehensive picture.

As to the type of judgments, we explicitly focused on rating scales instead of binary decisions. Binary decisions might be appropriate for clear-cut grammatical contrasts but even in those cases they may conceal meaningful information as, for instance, native speakers do not consider all ungrammatical sentences equally unacceptable (see Ionin & Zyzik, 2014 for a great review of judgment and interpretation tasks in LX research). And indeed, the results lend support to the idea that irony, sensibility and other judgments are not all-or-nothing phenomena. Importantly, however, Experiments 2 and 4 used a rather large scale (50-point and 100-point, respectively). Future research may directly investigate whether larger scales (e.g., 20-point or larger) provide more detailed results than regular 7-point Likert scales. One disadvantage of such larger scales is that ordinal generalized additive mixed models (GAMM; Divjak & Baayen, 2017) take an extremely long time to compute, hence one has to resort to GAMMs of the Gaussian family or linear mixed modelling when the data, in fact, is not integer-valued.

Finally, due to either external reasons such as COVID or internal reasons related to the design of the studies, this dissertation had a limited number of speakers per each accent (two). Although in Experiment 4 two different non-native accents were used (albeit from the same language family), Experiment 1 used only Mandarin Chinese. This creates a possibility that the effects obtained 1) have more to do with the idiosyncrasies of particular speakers than non-native speakers in general, or 2) are affected by the stereotypes associated with a particular language community. Future studies should recruit more speakers per accent to limit the impact of these effects or specifically compare different accents to each other.

6.4.4 Where to from here

This dissertation provided clear evidence that the identity of both the speaker and the listener/reader affects linguistic judgments during the resolution of pragmatic or epistemic uncertainties. Ironies were found to be rated as less ironic and both literal and metaphoric sentences as making less sense when the speaker was non-native. Truth evaluation was also affected by the speaker identity, although in a less straightforward way. Two questions thus need to be addressed by future research: how these effects are correlated with online processing and what they entail for non-native speakers behaviourally and evaluatively. I will discuss these two questions in more detail below.

First, this dissertation focused on cognitive judgments that only reflect the outcome of pragmatic processing. Therefore, future research needs to shed more light on how these judgments relate to online sentence processing. For instance, are speaker effects visible in reaction times during self-paced reading or listening, in ERP components, pupil dilation, or eye-tracking data? How do these time-sensitive methods correlate with explicit linguistic judgments? Whether individual differences found to affect judgments in this dissertation would affect online processing remains to be seen. In addition, as the experiment on metaphors reported here only investigated judgements on one scale (sensibility), future research may employ other scales as well as investigate qualitative data with open-ended questions to shed more light on the source of the effect reported here.

Second, as non-native speakers were found to generally be at a disadvantage as far as evaluation of their language goes, it would be interesting to see whether these judgments have immediate behavioral and evaluative consequences for the speakers. Previous research has made initial steps in this direction. For instance, pragmatic violations of the maxims of informativeness and relevance have been shown to affect speaker evaluations along the dimensions of warmth and competence (Beltrama & Papafragou, 2023). In addition, M. H. K. Ip & Papafragou (2023) showed that violations of informativeness are more readily forgiven in non-native compared to native speakers, and that the former are rated as less dishonest. Even more interesting is the question whether such evaluations would have direct behavioral consequences for non-native speakers. For instance, in non-anonymous course assignments where students employ metaphors, would essays by native speakers be perceived as making more sense and thus graded higher? Could perceived figurative language "failures" be used to justify broader cultural stereotypes? Would answers from

non-native speakers during trivia pub quizzes be dismissed more readily? Since misunderstood ironic intents may injure the conversation and impede connecting with others on an emotional level, would mere communication with non-native speakers be rated as less engaging/fulfilling/satisfying when it is, in fact, native speakers who misinterpret the message in the first place? And how does all that interact with overt markers of language proficiency, such as syntactic and segmental/suprasegmental errors? All these questions have tangible consequences for billions of LX speakers around the world, who already face numerous challenges in day-to-day communication, which eventually translates into negative consequences for many aspects of life (Birney et al., 2020; Giles & Billings, 2004; Spence et al., 2024).

In addition, it is not clear how early in the course of one's life biases reported in this dissertation emerge. While it has been shown that infants prefer to look at native speakers and that kids more readily befriend native speakers than non-native speakers (Imuta & Spence, 2020; Kinzler et al., 2007, 2009; Nazzi et al., 2000), it remains to be seen whether young children make different pragmatic inferences from native and non-native speech and how early during their language acquisition the non-native speaker's identity is taken into account. Considering that children become sensitive to gender stereotypes at a young age (Canessa-Pollard et al., 2022), it would not be surprising if children responded differently to semantic anomalies, syntactic errors, or pragmatic violations when they are talking to non-native as compared to native speakers. All these avenues appear fruitful for future research and will significantly enhance our models of language comprehension.

Moreover, while all the above consequences are about potentially discriminatory behaviours against non-native speakers, it would be interesting to know whether non-native speakers *themselves* adjust their behaviour, either knowing about or merely suspecting possible communication breakdowns that may occur. Our acoustic analysis suggested that non-native speakers may engage in compensatory behaviours wherein they overly mark ironic statements prosodically to (potentially) avoid being misunderstood. It is possible that this behaviour extends to other linguistic domains as well, with non-native speakers overexplaining their intended message, more often labeling irony explicitly in digital communication (using emojis, capital letters, hashtags like #irony; see Mikhailova, 2020 for a review of some graphical markers of irony

online), sticking to more direct and literal tropes even when they are capable of more, or using overly formal language.

6.5 Conclusions

The studies reported in this dissertation provide consistent evidence that non-native speaker identity affects language comprehension but contradict the idea that only the language associated with interpretative uncertainty is affected. In fact, both literal and figurative statements, as well as facts with both known and unknown truth values, were sometimes equally affected depending on the set of stimuli and the task. Furthermore, resolution of linguistic versus epistemic uncertainty differed in speaker effects, with the former being more straightforwardly hindered by non-nativeness. Specifically, both types of ironies were rated as less ironic when they came from non-native compared to native speakers, whereas literal language remained largely unaffected by the accent. In contrast, both literal and metaphorical sentences were judged less sensical when merely attributed to a non-native speaker, thus the effect was not trope-specific. The frequency of figurative language tropes did emerge as a mediating factor. Rarer forms of figurative language were more affected by interpersonal variation and thus showed less pronounced speaker effects, suggesting that the effects of non-native accents may be more salient in-between linguistic extremes. Truth evaluation of domain-specific unknown statements may have been impaired by non-nativeness in the written modality only, but this effect was overridden by modality-specific social category change. Coupled with explicit signs of societal integration, a non-native accent was, in fact, beneficial for veracity judgments in the auditory modality. Furthermore, the resolution of linguistic and, to a smaller extent, epistemic ambiguity showed significant sensitivity to individual differences, including cognitive and socio-psychological traits. Notably, a strong implicit knowledgeability bias in favour of native-accented speakers and against foreign-accented speakers emerged in the Implicit Association Test, which was found to modulate truth evaluation in the auditory modality. This establishes generalized implicit accent bias as a viable predictor for language processing, motivating further exploration of this measure across different accents, populations, and communicative contexts. All in all, the findings of this dissertation suggest that the effects of non-nativeness dynamically interact with social and personal factors and underscore the importance of stimulus modality.

References

Abraham, A., Rutter, B., & Hermann, C. (2021). Conceptual expansion via novel metaphor processing: An ERP replication and extension study examining individual differences in creativity. *Brain and Language*, *221*, 105007. https://doi.org/10.1016/j.bandl.2021.105007

Ackerman, B. P. (1983). Form and function in children's understanding of ironic utterances. *Journal of Experimental Child Psychology*, *35*(3), 487–508. https://doi.org/10.1016/0022-0965(83)90023-1

Adank, P., Nuttall, H. E., Banks, B., & Kennedy-Higgins, D. (2015). Neural bases of accented speech perception. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00558

Adolphs, R., & Damasio, A. R. (2012). The interaction of affect and cognition: A neurobiological perspective. In *Handbook of affect and social cognition* (pp. 27–49). Psychology Press.

Al-Azary, H., & Buchanan, L. (2017). Novel metaphor comprehension: Semantic neighbourhood density interacts with concreteness. *Memory & Cognition*, *45*(2), 296–307. https://doi.org/10.3758/s13421-016-0650-7

Altemeyer, B. (1996). The authoritarian specter. *Harvard University*.

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the Tribes of Fluency to Form a Metacognitive Nation. *Personality and Social Psychology Review*, *13*(3), 219–235. https://doi.org/10.1177/1088868309341564

Amodio, D. M. (2014). The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience*, *15*(10), 670–682. https://doi.org/10.1038/nrn3800

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The Relationship Between Native Speaker Judgments of Nonnative Pronunciation and Deviance in Segmentais, Prosody, and Syllable Structure. *Language Learning*, *42*(4), 529–555. https://doi.org/10.1111/j.1467-1770.1992.tb01043.x

Anolli, L., Ciceri, R., & Infantino, M. G. (2000). Irony as a Game of Implicitness: Acoustic Profiles of Ironic Communication. *Journal of Psycholinguistic Research*, *29*(3), 37.

Anolli, L., Ciceri, R., & Infantino, M. G. (2002). From "blame by praise" to "praise by blame": Analysis of vocal patterns in ironic communication. *International Journal of Psychology*, *37*(5), 266–276. https://doi.org/10.1080/00207590244000106

Arppe, A., & Järvikivi, J. (2007). Take empiricism seriously! In support of methodological diversity in linguistics. *Corpus Linguistics and Linguistic Theory*, *3*(1), 99–109. https://doi.org/10.1515/CLLT.2006.007

Arzouan, Y., Goldstein, A., & Faust, M. (2007). Brainwaves are stethoscopes: ERP correlates of novel metaphor comprehension. *Brain Research*, *1160*, 69–81. https://doi.org/10.1016/j.brainres.2007.05.034

Attardo, S. (2000). Irony markers and functions: Toward a goal-oriented theory of irony and its processing. *Rask*, *12*(1), 3–20.

Attardo, S. (2001). *Humorous texts: A semantic and pragmatic analysis* (Vol. 6). Walter de Gruyter.

Attardo, S., Eisterhold, J., Hay, J., & Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor - International Journal of Humor Research*, *16*(2), 18. https://doi.org/10.1515/humr.2003.012

Averbeck, J. M. (2015). Irony in interpersonal conflict scenarios: Differences in ironic argument endorsement and suppression across topics. *Journal of Argumentation in Context*, *4*(1), 87–109. https://doi.org/10.1075/jaic.4.1.05ave

Averbeck, J. M., & Hample, D. (2008). Ironic Message Production: How and Why We Produce Ironic Messages. *Communication Monographs*, *75*(4), 396–410. https://doi.org/10.1080/03637750802512389

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, *39*(4), 437–456. https://doi.org/10.1017/S0047404510000400

Bambini, V., Bertini, C., Schaeken, W., Stella, A., & Di Russo, F. (2016). Disentangling Metaphor from Context: An ERP Study. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00559

Banton, O., West, K., & Kinney, E. (2020). The surprising politics of anti-immigrant prejudice: How political conservatism moderates the effect of immigrant race and religion on infrahumanization judgements. *British Journal of Social Psychology*, *59*(1), 157–170. https://doi.org/10.1111/bjso.12337

Barenboym, D. A., Wurm, L. H., & Cano, A. (2010). A comparison of stimulus ratings made online and in person: Gender and method effects. *Behavior Research Methods*, *42*(1), 273–285. https://doi.org/10.3758/BRM.42.1.273

Barlow, S., Beardsley, G., Bsharah, Z., Crofts, R., De La Rosa, C., Gutierrez, A., Highfill, C., Johnson, A. G. W., Johnson, C., Johnson, J., Cardenas, I. L., Martinez, J. T., Miller, N. T., Murray, R. M., Page, S., Petersen, T., Ramos, I., Rhodes, R., Vainuku, P., … Hayes-Harb, R. (2024). The effects of exposure and explicit stereotypes on veracity judgments of Polish-accented English speech: A preregistered close replication and extension of Boduch-Grabka & Lev-Ari (2021). *Studies in Second Language Acquisition*, *46*(5), 1320–1336. https://doi.org/10.1017/S0272263124000123

Bašnáková, J., Weber, K., Petersson, K. M., van Berkum, J., & Hagoort, P. (2014). Beyond the Language Given: The Neural Correlates of Inferring Speaker Meaning. *Cerebral Cortex*, *24*(10), 2572–2578. https://doi.org/10.1093/cercor/bht112

Bates, D., Maechler, M., Bolker, B., Walker, S., Haubo Bojesen Christensen, R., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., & Krivitsky, P. N. (2023). *lme4: Linear Mixed-Effects Models using "Eigen" and S4* (Version 1.1-35.1) [Computer software]. https://cran.r-project.org/web/packages/lme4/index.html

Bazzi, L., Brouwer, S., & Foucart, A. (2022). The impact of foreign accent on irony and its consequences on social interaction. *Journal of Multilingual and Multicultural Development*, 1–13. https://doi.org/10.1080/01434632.2022.2126487

Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, *121*(4), 446.

Beltrama, A. (2020). Social meaning in semantics and pragmatics. *Language and Linguistics Compass*, *14*(9), e12398. https://doi.org/10.1111/lnc3.12398

Beltrama, A., & Papafragou, A. (2023). Pragmatic violations affect social inferences about the speaker. *Glossa Psycholinguistics*, *2*(1). https://doi.org/10.5070/G601197

Beltrama, A., Solt, S., & Burnett, H. (2023). Context, precision, and social perception: A sociopragmatic study. *Language in Society*, *52*(5), 805–835. https://doi.org/10.1017/S0047404522000240

Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1450–1460. https://doi.org/10.1037/a0027850

Birdsong, D. (2006). Age and Second Language Acquisition and Processing: A Selective Overview. *Language Learning*, *56*(s1), 9–49. https://doi.org/10.1111/j.1467-9922.2006.00353.x

Birney, M. E., Rabinovich, A., Morton, T. A., Heath, H., & Ashcroft, S. (2020). When Speaking English Is Not Enough: The Consequences of Language-Based Stigma for Nonnative Speakers. *Journal of Language and Social Psychology*, *39*(1), 67–86. https://doi.org/10.1177/0261927X19883906

Blasko, D. G., & Connine, C. M. (1993). Effects of Familiarity and Aptness on Metaphor Processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 295–308.

Block, J., & Block, J. H. (2006). Nursery school personality and political orientation two decades later. *Journal of Research in Personality*, *40*(5), 734–749. https://doi.org/10.1016/j.jrp.2005.09.005

Boduch-Grabka, K., & Lev-Ari, S. (2021). Exposing individuals to foreign accent increases their trust in what nonnative speakers say. *Cognitive Science*, *45*(11), e13064.

Boersma, P., & Van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glot International*, *5*(9/10), 341–347.

Bonett, D. G. (2002). Sample Size Requirements for Testing and Estimating Coefficient Alpha. *Journal of Educational and Behavioral Statistics*, *27*(4), 335–340. https://doi.org/10.3102/10769986027004335

Bornkessel-Schlesewsky, I., Krauspenhaar, S., & Schlesewsky, M. (2013). Yes, You Can? A Speaker's Potency to Act upon His Words Orchestrates Early Neural Responses to Message-Level Meaning. *PLoS ONE*, *8*(7), e69173. https://doi.org/10.1371/journal.pone.0069173

Boyd, R. N. (1993). *Metaphor and theory change*.

Bradac, J. J., Courtright, J. A., Schmidt, G., & Davies, R. A. (1976). The effects of perceived status and linguistic diversity upon judgments of speaker attributes and message effectiveness. *The Journal of Psychology*, *93*(2), 213–220.

Bradac, J. J., & Mulac, A. (1984). Attributional Consequences of Powerful and Powerless Speech Styles in a Crisis-Intervention Context. *Journal of Language and Social Psychology*, *3*(1), 1–19. https://doi.org/10.1177/0261927X8431001

Bradac, J. J., & Wisegarver, R. (1984). Ascribed Status, Lexical Diversity, and Accent: Determinants of Perceived Status, Solidarity, and Control of Speech Style. *Journal of Language and Social Psychology*, *3*(4), 239–255. https://doi.org/10.1177/0261927X8400300401

Breen, M. (2014). Empirical Investigations of the Role of Implicit Prosody in Sentence Processing. *Language and Linguistics Compass*, *8*(2), 37–50. https://doi.org/10.1111/lnc3.12061

Bresnahan, M. J., Ohashi, R., Nebashi, R., Liu, W. Y., & Morinaga Shearman, S. (2002). Attitudinal and affective response toward accented English. *Language & Communication*, *22*(2), 171–185. https://doi.org/10.1016/S0271-5309(01)00025-8

Bromberek-Dyzman, K., & Rataj, K. (2016). Irony Comprehension in the Nonnative Language Comes at a Cost. *Psychology of Language and Communication*, *20*(3), 336–353. https://doi.org/10.1515/plc-2016-0020

Brown, P., & Levinson, S. C. (1987). *Politeness: Some Universals in Language Usage* (Vol. 4). Cambridge university press.

Bruntsch, R., Hofmann, J., & Ruch, W. (2016). Virgin soil in irony research: Personality, humor, and the "sense of irony". *Translational Issues in Psychological Science*, *2*(1), 25–34. https://doi.org/10.1037/tps0000054

Bruntsch, R., & Ruch, W. (2017). Studying irony detection beyond ironic criticism: Let's include ironic praise. *Frontiers in Psychology*, *8*, 606. https://doi.org/10.3389/fpsyg.2017.00606

Bryant, G. A. (2010). Prosodic Contrasts in Ironic Speech. *Discourse Processes*, *47*(7), 545–566. https://doi.org/10.1080/01638530903531972

Bryant, G. A. (2012). Is Verbal Irony Special? *Language and Linguistics Compass*, *6*(11), 673–685. https://doi.org/10.1002/lnc3.364

Bryant, G. A., & Fox Tree, J. E. (2005). Is there an Ironic Tone of Voice? *Language and Speech*, *48*(3), 257–277. https://doi.org/10.1177/00238309050480030101

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Caffarra, S., & Martin, C. D. (2019). Not all errors are the same: ERP sensitivity to error typicality in foreign accented speech perception. *Cortex*, *116*, 308–320. https://doi.org/10.1016/j.cortex.2018.03.007

Caffarra, S., Michell, E., & Martin, C. D. (2018). The impact of foreign accent on irony interpretation. *PLOS ONE*, *13*(8), e0200939. https://doi.org/10.1371/journal.pone.0200939

Caffarra, S., Motamed Haeri, A., Michell, E., & Martin, C. D. (2019). When is irony influenced by communicative constraints? ERP evidence supporting interactive models. *European Journal of Neuroscience*, *50*(10), 3566–3577. https://doi.org/10.1111/ejn.14503

Caillies, S., Gobin, P., Obert, A., Terrien, S., Coutté, A., Iakimova, G., & Besche-Richard, C. (2019). Asymmetry of affect in verbal irony understanding: What about the N400 and P600

components? *Journal of Neurolinguistics*, *51*, 268–277. https://doi.org/10.1016/j.jneuroling.2019.04.004

Calkins, K., Puhacheuskaya, V., Kim, J., & Järvikivi, J. (2023). Labels speak volumes: The role of intergroup membership in accent perception. *Proceedings of the 20th International Congress of Phonetic Sciences*.

Canessa-Pollard, V., Reby, D., Banerjee, R., Oakhill, J., & Garnham, A. (2022). The development of explicit occupational gender stereotypes in children: Comparing perceived gender ratios and competence beliefs. *Journal of Vocational Behavior*, *134*, 103703. https://doi.org/10.1016/j.jvb.2022.103703

Cardillo, E. R., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2010). Stimulus design is an obstacle course: 560 matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behavior Research Methods*, *42*(3), 651–664. https://doi.org/10.3758/BRM.42.3.651

Cardillo, E. R., Watson, C. E., Schmidt, G. L., Kranjec, A., & Chatterjee, A. (2012). From novel to familiar: Tuning the brain for metaphors. *NeuroImage*, *59*(4), 3212–3221. https://doi.org/10.1016/j.neuroimage.2011.11.079

Cargile, A. C., & Bradac, J. J. (2001). Attitudes Toward Language: A Review of Speaker-Evaluation Research and a General Process Model. *Annals of the International Communication Association*, *25*(1), 347–382. https://doi.org/10.1080/23808985.2001.11679008

Cargile, A. C., Giles, H., Ryan, E. B., & Bradac, J. J. (1994). Language attitudes as a social process: A conceptual model and new directions. *Language & Communication*, *14*(3), 211–236. https://doi.org/10.1016/0271-5309(94)90001-9

Carraro, L., Castelli, L., & Macchiella, C. (2011). The Automatic Conservative: Ideology-Based Attentional Asymmetries in the Processing of Valenced Information. *PLoS ONE*, *6*(11), e26456. https://doi.org/10.1371/journal.pone.0026456

Carraro, L., Negri, P., Castelli, L., & Pastore, M. (2014). Implicit and Explicit Illusory Correlation as a Function of Political Ideology. *PLoS ONE*, *9*(5), e96312. https://doi.org/10.1371/journal.pone.0096312

Carroll, M., & Korukina, S. (1999). The Effect of Text Coherence and Modality on Metamemory Judgements. *Memory*, *7*(3), 309–322. https://doi.org/10.1080/096582199387940

Castelli, L., & Carraro, L. (2011). Ideology is related to basic cognitive processes involved in attitude formation. *Journal of Experimental Social Psychology*, *47*(5), 1013–1016. https://doi.org/10.1016/j.jesp.2011.03.016

Champoux-Larsson, M.-F., Ramström, F., Costa, A., & Baus, C. (2022). Social Categorization Based on Language and Facial Recognition. *Journal of Language and Social Psychology*, *41*(3), 331–349. https://doi.org/10.1177/0261927X211035159

Cheang, H. S., & Pell, M. D. (2008). The sound of sarcasm. *Speech Communication*, *50*(5), 366–381. https://doi.org/10.1016/j.specom.2007.11.003

Cheng, L. S. P., Burgess, D., Vernooij, N., Solís-Barroso, C., McDermott, A., & Namboodiripad, S. (2021). The Problematic Concept of Native Speaker in Psycholinguistics: Replacing Vague and Harmful Terminology With Inclusive and Accurate Measures. *Frontiers in Psychology*, *12*, 715843. https://doi.org/10.3389/fpsyg.2021.715843

Chevance, G., Héraud, N., Guerrieri, A., Rebar, A., & Boiché, J. (2017). Measuring implicit attitudes toward physical activity and sedentary behaviors: Test-retest reliability of three scoring algorithms of the Implicit Association Test and Single Category-Implicit Association Test. *Psychology of Sport and Exercise*, *31*, 70–78. https://doi.org/10.1016/j.psychsport.2017.04.007

Chirumbolo, A., Areni, A., & Sensales, G. (2004). Need for cognitive closure and politics: Voting, political attitudes and attributional style. *International Journal of Psychology*, *39*(4), 245–253. https://doi.org/10.1080/00207590444000005

Choma, B. L., & Hodson, G. (2008). *And so the pendulum swings: A framework for conceptualizing the causes of prejudice.*

Chomsky, N. (1965). *Aspects of the Theory of Syntax* (Issue 11). MIT press.

Cichocka, A., Bilewicz, M., Jost, J. T., Marrouch, N., & Witkowska, M. (2016). On the Grammar of Politics-or Why Conservatives Prefer Nouns: On the Grammar of Politics. *Political Psychology*, *37*(6), 799–815. https://doi.org/10.1111/pops.12327

Cieślicka, A. B. (2015). Idiom acquisition and processing by second/foreign language learners. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual Figurative Language Processing* (pp. 208–244). Cambridge University Press. https://doi.org/10.1017/CBO9781139342100.012

Clark, H. H., & Gerrig, R. J. (1984). On the pretense theory of irony. *Journal of Experimental Psychology: General*, *113*(1), 121–126. https://doi.org/10.1037/0096-3445.113.1.121

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented Englisha). *J. Acoust. Soc. Am.*, *116*(6).

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407.

Collins, K. A., & Clément, R. (2012). Language and Prejudice: Direct and Moderated Effects. *Journal of Language and Social Psychology*, *31*(4), 376–396. https://doi.org/10.1177/0261927X12446611

Colston, H. L., & Gibbs, R. W. (2002). Are irony and metaphor understood differently? *Metaphor and Symbol*, *17*(1), 57–80. https://doi.org/10.1207/S15327868MS1701_5

Contemori, C., & Tortajada, F. (2020). The use of social–communicative cues to interpret ambiguous pronouns: Bilingual adults differ from monolinguals. *Applied Psycholinguistics*, *41*(1), 51–77. https://doi.org/10.1017/S0142716419000407

Converse, B. A., Lin, S., Keysar, B., & Epley, N. (2008). In the mood to get over yourself: Mood affects theory-of-mind use. *Emotion*, *8*(5), 725–730. https://doi.org/10.1037/a0013283

Converse, P. E. (1964). The nature of belief systems in mass publics. *Critical Review*, *18*(1–3), 1–74. https://doi.org/10.1080/08913810608443650

Conway, M. A., & Gathercole, S. E. (1987). Modality and long-term memory. *Journal of Memory and Language*, *26*(3), 341–361.

Cook, V. (2002). Portraits of the L2 user. *Multilingual Matters*.

Costello, T. H., Bowes, S. M., Stevens, S. T., Waldman, I., Tasimi, A., & Lilienfeld, S. O. (2020). *Clarifying the Structure and Nature of Left-Wing Authoritarianism*. https://doi.org/10.31234/osf.io/3nprq

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*(4), 349.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, *92*(4), 631–648. https://doi.org/10.1037/0022-3514.92.4.631

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map. In *Advances in Experimental Social Psychology* (Vol. 40, pp. 61–149). Elsevier. https://doi.org/10.1016/S0065-2601(07)00002-0

Cutler, A. (1974). On saying what you mean without meaning what you say. *Papers from the Tenth Regional Meeting*, 117–127.

Cutler, A., & Clifton, C. (1999). Comprehending spoken language: A blueprint of the listener. *The Neurocognition of Language*, 123–166.

Dasgupta, N., & Greenwald, A. G. (2001). On the Malleability of Automatic Attitudes: Combating Automatic Prejudice With Images of Admired and Disliked Individuals. *Journal of Personality and Social Psychology*, *81*(5), 800–814. https://doi.org/10.1037/0022-3514.81.5.800

Davies, A. (2003). *The native speaker: Myth and reality* (2nd ed). Multilingual Matters.

Davies, A. (2004). The native speaker in applied linguistics. *The Handbook of Applied Linguistics*, 431–450.

Davis, M. H. (1980). *A multidimensional approach to individual differences in empathy*.

de Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390. https://doi.org/10.3758/BRM.41.2.385

Dennis, M., Purvis, K., Barnes, M. A., Wilkinson, M., & Winner, E. (2001). Understanding of literal truth, ironic criticism, and deceptive praise following childhood head injury. *Brain and Language*, *78*(1), 1–16. https://doi.org/10.1006/brln.2000.2431

Derwing, T. (2003). What do ESL students say about their accents? *Canadian Modern Language Review*, *59*(4), 547–567. https://doi.org/10.3138/cmlr.59.4.547

Dewaele, J.-M. (2017). Why the Dichotomy 'L1 Versus LX User' is Better than 'Native Versus Non-native Speaker. *Applied Linguistics*, amw055. https://doi.org/10.1093/applin/amw055

Dewaele, J.-M., & McCloskey, J. (2015). Attitudes toward foreign accents among adult multilingual language users. *Journal of Multilingual and Multicultural Development*, *36*(3), 221–238. https://doi.org/10.1080/01434632.2014.909445

Dews, S., Kaplan, J., & Winner, E. (1995). Why not say it directly? The social functions of irony. *Discourse Processes*, *19*(3), 347–367. https://doi.org/10.1080/01638539509544922

Dews, S., Winner, E., Nicolaides, N., & Hunt, M. (1995). Forms and functions of verbal irony found in children's and adults' television shows. *American Psychological Association, New York*.

DeYoung, C. G., & Gray, J. R. (2009). Personality neuroscience: Explaining individual differences in affect, behaviour and cognition. In P. J. Corr & G. Matthews (Eds.), *The Cambridge Handbook of Personality Psychology* (pp. 323–346). Cambridge University Press. https://doi.org/10.1017/CBO9780511596544.023

Diehl, M. (1990). The Minimal Group Paradigm: Theoretical Explanations and Empirical Findings. *European Review of Social Psychology*, *1*(1), 263–292. https://doi.org/10.1080/14792779108401864

Divjak, D., & Baayen, H. (2017). Ordinal GAMMs: A new window on human ratings. In *Each venture, a new beginning: Studies in Honor of Laura A. Janda* (pp. 39–56). Slavica Publishers.

Dollmann, J., Kogan, I., & Weißmann, M. (2020). Speaking Accent-Free in L2 Beyond the Critical Period: The Compensatory Role of Individual Abilities and Opportunity Structures. *Applied Linguistics*, *41*(5), 787–809. https://doi.org/10.1093/applin/amz029

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLOS ONE*, *18*(3), e0279720. https://doi.org/10.1371/journal.pone.0279720

Dovidio, J. F., & Gaertner, S. L. (2010). Intergroup Bias. In *Handbook of Social Psychology* (In S. T. Fiske, D. T. Gilbert, G. Lindzey (Eds.), pp. 1084–1121). John Wiley & Sons, Inc. 10.1002/9780470561119.socpsy002029

Dragojevic, M., & Giles, H. (2016). I Don't Like You Because You're Hard to Understand: The Role of Processing Fluency in the Language Attitudes Process: Processing Fluency and Language Attitudes. *Human Communication Research*, *42*(3), 396–420. https://doi.org/10.1111/hcre.12079

Dragojevic, M., Giles, H., Beck, A.-C., & Tatum, N. T. (2017). The fluency principle: Why foreign accent strength negatively biases language attitudes. *Communication Monographs*, *84*(3), 385–405. https://doi.org/10.1080/03637751.2017.1322213

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2021). *Ethnologue: Languages of the World.* SIL International. http://www.ethnologue.com.

Eckert, P., & Labov, W. (2017). Phonetics, phonology and social meaning. *Journal of Sociolinguistics*, *21*(4), 467–496. https://doi.org/10.1111/josl.12244

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.*

Eekhof, L. S., van Krieken, K., Sanders, J., & Willems, R. M. (2021). Reading Minds, Reading Stories: Social-Cognitive Abilities Affect the Linguistic Processing of Narrative Viewpoint. *Frontiers in Psychology*, *12*, 698986. https://doi.org/10.3389/fpsyg.2021.698986

Egidi, G., & Caramazza, A. (2014). Mood-dependent integration in discourse comprehension: Happy and sad moods affect consistency processing via different brain networks. *NeuroImage*, *103*, 20–32. https://doi.org/10.1016/j.neuroimage.2014.09.008

Elder, C.-H., & Haugh, M. (2024). The role of inference and inferencing in pragmatic models of communication. *Journal of Pragmatics*, *229*, 71–76. https://doi.org/10.1016/j.pragma.2024.06.002

Ellis, R., Zhu, Y., Shintani, N., & Roever, C. (2021). A study of Chinese learners' ability to comprehend irony. *Journal of Pragmatics*, *172*, 7–20. https://doi.org/10.1016/j.pragma.2020.11.003

Escandell-Vidal, V., & Prieto, P. (2020). Pragmatics and prosody in research on Spanish. In *The Routledge Handbook of Spanish Pragmatics* (1st Edition, p. 18). Routledge.

Evans, J. R., & Michael, S. W. (2014). Detecting Deception in Non-Native English Speakers. *Applied Cognitive Psychology*, *28*(2), 226–237. https://doi.org/10.1002/acp.2990

Fairchild, S., Mathis, A., & Papafragou, A. (2020). Pragmatics and social meaning: Understanding under-informativeness in native and non-native speakers. *Cognition*, *200*, 104171. https://doi.org/10.1016/j.cognition.2019.104171

Fairchild, S., & Papafragou, A. (2018). Sins of omission are more likely to be forgiven in non-native speakers. *Cognition*, *181*, 80–92. https://doi.org/10.1016/j.cognition.2018.08.010

Fasiolo, M., Nedellec, R., Goude, Y., Capezza, C., & Wood, N. S. (2023). *mgcViz: Visualisations for Generalized Additive Models* (Version 0.1.11) [Computer software]. https://cran.r-project.org/web/packages/mgcViz/index.html

Federico, C. M., & Goren, P. (2009). Motivated social cognition and ideology: Is attention to elite discourse a prerequisite for epistemically motivated political affinities. *Social and Psychological Bases of Ideology and System Justification*, 267–291.

Federmeier, K. D., Kirson, D. A., Moreno, E. M., & Kutas, M. (2001). Effects of transient, mild mood states on semantic memory organization and use: An event-related potential investigation in humans. *Neuroscience Letters*, *305*(3), 149–152. https://doi.org/10.1016/S0304-3940(01)01843-2

Feldman, S. (1988). Structure and Consistency in Public Opinion: The Role of Core Beliefs and Values. *American Journal of Political Science*, *32*(2), 416. https://doi.org/10.2307/2111130

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Flege, J. E. (1984). The detection of French accent by American listeners. *The Journal of the Acoustical Society of America*, *76*(3), 692–707.

Fodor, J. A. (1983). *The modularity of mind*. MIT press.

Foucart, A., Costa, A., Morís-Fernández, L., & Hartsuiker, R. J. (2020). Foreignness or Processing Fluency? On Understanding the Negative Bias Toward Foreign-Accented Speakers. *Language Learning*, *70*(4), 974–1016. https://doi.org/10.1111/lang.12413

Foucart, A., & Hartsuiker, R. J. (2021). Are foreign-accented speakers that 'incredible'? The impact of the speaker's indexical properties on sentence processing. *Neuropsychologia*, *158*, 107902. https://doi.org/10.1016/j.neuropsychologia.2021.107902

Foucart, A., Santamaría-García, H., & Hartsuiker, R. J. (2019). Short exposure to a foreign accent impacts subsequent cognitive processes. *Neuropsychologia*, *129*, 1–9. https://doi.org/10.1016/j.neuropsychologia.2019.02.021

Fox, J., & Weisberg, S. (2018). *An R companion to applied regression*. Sage publications.

Frances, C., Costa, A., & Baus, C. (2018). On the effects of regional accents on memory and credibility. *Acta Psychologica*, *186*, 63–70. https://doi.org/10.1016/j.actpsy.2018.04.003

Fraser, C., & Kelly, B. F. (2012). Listening between the lines: Social assumptions around foreign accents. *Australian Review of Applied Linguistics*, *35*(1), 74–93. https://doi.org/10.1075/aral.35.1.04fra

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Frenkel-Brunswik, E. (1949). Intolerance of ambiguity as an emotional and perceptual personality variable. *Journal of Personality*, *18*(1), 108–143. https://doi.org/10.1111/j.1467-6494.1949.tb01236.x

Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, *16*(5), 262–268. https://doi.org/10.1016/j.tics.2012.04.001

Friederici, A. D. (2017). *Language in our brain: The origins of a uniquely human capacity*. The MIT Press.

Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C., & Giles, H. (2012). A meta-analysis of the effects of speakers' accents on interpersonal evaluations: Effects of speakers' accents. *European Journal of Social Psychology*, *42*(1), 120–133. https://doi.org/10.1002/ejsp.862

Gaertner, S. L., Dovidio, J. F., Anastasio, P. A., Bachman, B. A., & Rust, M. C. (1993). The Common Ingroup Identity Model: Recategorization and the Reduction of Intergroup Bias. *European Review of Social Psychology*, *4*(1), 1–26. https://doi.org/10.1080/14792779343000004

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731. https://doi.org/10.1037/0033-2909.132.5.692

Gibbs Jr., R. W. (2000). Irony in Talk Among Friends. *Metaphor and Symbol*, *15*(1–2), 5–27. https://doi.org/10.1080/10926488.2000.9678862

Gibbs Jr., R. W. (2002). A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics*, *34*(4), 457–486. https://doi.org/10.1016/S0378-2166(01)00046-7

Gibbs, R. W. (1984). Literal Meaning and Psychological Theory. *Cognitive Science*, *8*(3), 275–304. https://doi.org/10.1207/s15516709cog0803_4

Gibbs, R. W. (1986). On the psycholinguistics of sarcasm. *Journal of Experimental Psychology: General*, *115*(1), 3.

Gibbs, R. W. (2000). Irony in Talk Among Friends. *Metaphor and Symbol*, *15*(1–2), 5–27. https://doi.org/10.1080/10926488.2000.9678862

Gibbs, R. W., & Colston, H. L. (2012). *Interpreting figurative meaning*. Cambridge University Press.

Gibbs, R. W., Kushner, J. M., & Mills, W. R. (1991). Authorial intentions and metaphor comprehension. *Journal of Psycholinguistic Research*, *20*(1), 11–30. https://doi.org/10.1007/BF01076917

Gibbs, R. W., & Perlman, M. (2010). Language understanding is grounded in experiential simulations: A response to Weiskopf. *Studies in History and Philosophy of Science Part A*, *41*(3), 305–308. https://doi.org/10.1016/j.shpsa.2010.07.004

Gibson, E., Tan, C., Futrell, R., Mahowald, K., Konieczny, L., Hemforth, B., & Fedorenko, E. (2017). Don't Underestimate the Benefits of Being Misunderstood. *Psychological Science*, *28*(6), 703–712. https://doi.org/10.1177/0956797617690277

Giles, H. (1970). Evaluative reactions to accents. *Educational Review*, *22*(3), 211–227. https://doi.org/10.1080/0013191700220301

Giles, H. (1973). Accent Mobility: A Model and Some Data. *Anthropological Linguistics*, *15*(2), 87–105.

Giles, H., & Billings, A. C. (2004). Assessing language attitudes: Speaker evaluation studies. *The Handbook of Applied Linguistics*, 187–209.

Giles, H., Wilson, P., & Conway, A. (1981). Accent and lexical diversity as determinants of impression formation and perceived employment suitability. *Language Sciences*, *3*(1), 91–103.

Giora, R. (1999). On the priority of salient meanings: Studies of literal and figurative language. *Journal of Pragmatics*, *31*(7), 919–929. https://doi.org/10.1016/S0378-2166(98)00100-3

Giora, R. (2002). Literal vs. figurative language: Different or equal? *Journal of Pragmatics*, *34*(4), 487–506. https://doi.org/10.1016/S0378-2166(01)00045-5

Glashouwer, K. A., Smulders, F. T. Y., De Jong, P. J., Roefs, A., & Wiers, R. W. H. J. (2013). Measuring automatic associations: Validation of algorithms for the Implicit Association

Test (IAT) in a laboratory setting. *Journal of Behavior Therapy and Experimental Psychiatry*, *44*(1), 105–113. https://doi.org/10.1016/j.jbtep.2012.07.015

Gluszek, A., & Dovidio, J. F. (2010). Speaking with a nonnative accent: Perceptions of bias, communication difficulties, and belonging in the United States. *Journal of Language and Social Psychology*, *29*(2), 224–234. https://doi.org/10.1177/0261927X09359590

Gluszek, A., & Hansen, K. (2016). *What does speaking with a foreign accent mean? Content analysis of newspaper articles*. Unpublished. https://doi.org/10.13140/RG.2.1.1912.3281

González Fuente, S. (2017). *Audiovisual prosody and verbal irony*.

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*, *122*(2), 92–102. https://doi.org/10.1016/j.bandl.2012.04.017

Greene, J. (2014). *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. Penguin.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes. *Psychological Review*, *102*(1), 4–27.

Greenwald, A. G., & Banaji, M. R. (2017). The implicit revolution: Reconceiving the relation between conscious and unconscious. *American Psychologist*, *72*(9), 861–871. https://doi.org/10.1037/amp0000238

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity. *Journal of Personality and Social Psychology*, *97*(1), 17–41. https://doi.org/10.1037/a0015575

Grey, S., Cosgrove, A. L., & van Hell, J. G. (2020). Faces with foreign accents: An event-related potential study of accented sentence comprehension. *Neuropsychologia*, *147*, 107575. https://doi.org/10.1016/j.neuropsychologia.2020.107575

Grey, S., & van Hell, J. G. (2017). Foreign-accented speaker identity affects neural correlates of language comprehension. *Journal of Neurolinguistics*, *42*, 93–108. https://doi.org/10.1016/j.jneuroling.2016.12.001

Grice, H. P. (1975). Logic and conversation. In *Speech Acts* (pp. 41–58).

Gucman, M. (2016). The role of individual differences and situational factors in perception of verbal irony. *Psychology of Language and Communication*, *20*(3), 255–277. https://doi.org/10.1515/plc-2016-0016

Gut, U. (2012). The LeaP corpus: A multilingual corpus of spoken learner German and learner English. In *Multilingual corpora and multilingual corpus analysis* (pp. 3–23). John Benjamins.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of Word Meaning and World Knowledge in Language Comprehension. *Science*, *304*(5669), 438–441. https://doi.org/10.1126/science.1095455

Haiman, J. (1998). *Talk is cheap: Sarcasm, alienation, and the evolution of language*. Oxford University Press.

Hammond-Thrasher, S., & Järvikivi, J. (2023). Rating gender stereotype violations: The effects of personality and politics. *Frontiers in Communication*, *8*, 1050662. https://doi.org/10.3389/fcomm.2023.1050662

Han, H. A., Czellar, S., Olson, M. A., & Fazio, R. H. (2010). Malleability of attitudes or malleability of the IAT? *Journal of Experimental Social Psychology*, *46*(2), 286–298. https://doi.org/10.1016/j.jesp.2009.11.011

Hancock, J. T. (2004). Verbal Irony Use in Face-To-Face and Computer-Mediated Conversations. *Journal of Language and Social Psychology*, *23*(4), 447–463. https://doi.org/10.1177/0261927X04269587

Hancock, J. T., Dunham, P. J., & Purdy, K. (2000). Children's comprehension of critical and complimentary forms of verbal irony. *Journal of Cognition and Development*, *1*(2), 227–248. https://doi.org/10.1207/S15327647JCD010204

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887. https://doi.org/10.1162/jocn_a_00103

Hanulíková, A., & Weber, A. (2012). Sink positive: Linguistic experience with th substitutions influences nonnative word recognition. *Attention, Perception, & Psychophysics*, *74*(3), 613–629. https://doi.org/10.3758/s13414-011-0259-7

Hanzlíková, D., & Skarnitzl, R. (2017). Credibility of native and non-native speakers of English revisited: Do non-native listeners feel the same? *Research in Language*, *15*(3), 285–298. https://doi.org/10.1515/rela-2017-0016

Harris, M., & Pexman, P. M. (2003). Children's perceptions of the social functions of verbal irony. *Discourse Processes*, *36*(3), 147–165. https://doi.org/10.1207/S15326950DP3603_1

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*, *23*(3). https://doi.org/10.1515/TLR.2006.014

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*, 575–604.

Hinton, P. (2017). Implicit stereotypes and the predictive brain: Cognition and culture in "biased" person perception. *Palgrave Communications*, *3*(1), 17086. https://doi.org/10.1057/palcomms.2017.86

Hodson, G., & Dhont, K. (2015). The person-based nature of prejudice: Individual difference predictors of intergroup negativity. *European Review of Social Psychology*, *26*(1), 1–42. https://doi.org/10.1080/10463283.2015.1070018

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A Meta-Analysis on the Correlation Between the Implicit Association Test and Explicit Self-Report Measures. *Personality and Social Psychology Bulletin*, *31*(10), 1369–1385. https://doi.org/10.1177/0146167205275613

Hogg, M. A., Joyce, N., & Abrams, D. (1984). Diglossia in Switzerland? A Social Identity Analysis of Speaker Evaluations. *Journal of Language and Social Psychology*, *3*(3), 185–196. https://doi.org/10.1177/0261927X8400300302

Holliday, A. (2006). Native-speakerism. *ELT Journal*, *60*(4), 385–387. https://doi.org/10.1093/elt/ccl030

Holt, R., Kung, C., & Demuth, K. (2018). Listener characteristics modulate the semantic processing of native vs. Foreign-accented speech. *PLOS ONE*, *13*(12), e0207452. https://doi.org/10.1371/journal.pone.0207452

Holyoak, K. J., & Stamenković, D. (2018). Metaphor comprehension: A critical review of theories and evidence. *Psychological Bulletin*, *144*(6), 641.

Hubert Lyall, I. (2019). *It's personal and disgusting: Extra-linguistic information in language comprehension*. University of Alberta.

Hubert Lyall, I., & Järvikivi, J. (2021). Individual Differences in Political Ideology and Disgust Sensitivity Affect Real-Time Spoken Language Comprehension. *Frontiers in Psychology*, *12*, 699071. https://doi.org/10.3389/fpsyg.2021.699071

Hymes, D. (1972). On communicative competence. *Sociolinguistics/Penguin*.

Imuta, K., & Spence, J. L. (2020). Developments in the social meaning underlying accent- and dialect-based social preferences. *Child Development Perspectives*, *14*(3), 135–141. https://doi.org/10.1111/cdep.12374

Ingvalson, E. M., Lansford, K. L., Federova, V., & Fernandez, G. (2017). Listeners' attitudes toward accented talkers uniquely predicts accented speech perception. *The Journal of the Acoustical Society of America*, *141*(3), EL234–EL238. https://doi.org/10.1121/1.4977583

Ionin, T., & Zyzik, E. (2014). Judgment and Interpretation Tasks in Second Language Research. *Annual Review of Applied Linguistics*, *34*, 37–64. https://doi.org/10.1017/S0267190514000026

Ip, M. H. K., & Papafragou, A. (2023). The pragmatics of foreign accents: The social costs and benefits of being a non-native speaker. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(9), 1505–1521. https://doi.org/10.1037/xlm0001197

Ip, M., & Papafragou, A. (2021). Listeners evaluate native and non-native speakers differently (but not in the way you think). *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*.

Ito, T. A., & Bartholow, B. D. (2009). The neural correlates of race. *Trends in Cognitive Sciences*, *13*(12), 524–531. https://doi.org/10.1016/j.tics.2009.10.002

Ito, T. A., & Cacioppo, J. T. (2000). Electrophysiological Evidence of Implicit and Explicit Categorization Processes. *Journal of Experimental Social Psychology*, *36*(6), 660–676. https://doi.org/10.1006/jesp.2000.1430

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, *85*(4), 616–626. https://doi.org/10.1037/0022-3514.85.4.616

Jakimik, J., & Glenberg, A. (1990). Verbal learning meets psycholinguistics: Modality effects in the comprehension of anaphora. *Journal of Memory and Language*, *29*(5), 582–590.

Janse, E., & Adank, P. (2012). Predicting foreign-accent adaptation in older adults. *Quarterly Journal of Experimental Psychology*, *65*(8), 1563–1585. https://doi.org/10.1080/17470218.2012.658822

Jiang, X., Gossack-Keenan, K., & Pell, M. D. (2020). To believe or not to believe? How voice and accent information in speech alter listener impressions of trust. *Quarterly Journal of Experimental Psychology*, *73*(1), 55–79. https://doi.org/10.1177/1747021819865833

Jiang, X., & Pell, M. D. (2015). On how the brain decodes vocal cues about speaker confidence. *Cortex*, *66*, 9–34. https://doi.org/10.1016/j.cortex.2015.02.002

Jiménez-Ortega, L., Hinchcliffe, C., Muñoz, F., Hernández-Gutiérrez, D., Casado, P., Sánchez-García, J., & Martín-Loeches, M. (2022). Language comprehension may depend on who you are: How personality traits and social presence seemingly modulate syntactic processing. *Language, Cognition and Neuroscience*, *37*(3), 365–380. https://doi.org/10.1080/23273798.2021.1970201

John, O. P., & Srivastava, S. (1991). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In *L. A. Pervin & O. P. John (Eds.). Handbook of personality: Theory and research* (Vol. 2, pp. 102–138). Guilford Press.

Jost, J. T., & Amodio, D. M. (2012). Political ideology as motivated social cognition: Behavioral and neuroscientific evidence. *Motivation and Emotion*, *36*(1), 55–64. https://doi.org/10.1007/s11031-011-9260-7

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, *129*(3), 339–375. https://doi.org/10.1037/0033-2909.129.3.339

Jost, J. T., Napier, J. L., Thorisdottir, H., Gosling, S. D., Palfai, T. P., & Ostafin, B. (2007). Are Needs to Manage Uncertainty and Threat Associated With Political Conservatism or Ideological Extremity? *Personality and Social Psychology Bulletin*, *33*(7), 989–1007. https://doi.org/10.1177/0146167207301028

Kamhi-Stein, L. D. (2016). The non-native English speaker teachers in TESOL movement. *ELT Journal*, *70*(2), 180–189. https://doi.org/10.1093/elt/ccv076

Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition*, *124*(1), 66–71. https://doi.org/10.1016/j.cognition.2012.03.001

Kanai, R., Feilden, T., Firth, C., & Rees, G. (2011). Political Orientations Are Correlated with Brain Structure in Young Adults. *Current Biology*, *21*(8), 677–680. https://doi.org/10.1016/j.cub.2011.03.017

Katz, A. N., Paivio, A., Marschark, M., & Clark, J. M. (1988). Norms for 204 literary and 260 nonliterary metaphors on 10 psychological dimensions. *Metaphor and Symbol*, *3*(4), 191–214.

Kidd, E. (2012). Individual differences in syntactic priming in language acquisition. *Applied Psycholinguistics*, *33*(2), 393–418. https://doi.org/10.1017/S0142716411000415

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

Kieckhäfer, C., Felsenheimer, A. K., & Rapp, A. M. (2019). A new test for irony detection: The influence of schizotypal, borderline, and autistic personality traits. *Frontiers in Psychiatry*, *10*, 28. https://doi.org/10.3389/fpsyt.2019.00028

Kiehne, E., & Ayón, C. (2016). Friends or Foes: The Impact of Political Ideology and Immigrant Friends on Anti-Immigrant Sentiment. *Journal of Sociology and Social Welfare*, *43*(3).

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, *104*(30), 12577–12580. https://doi.org/10.1073/pnas.0705345104

Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition*, *27*(4), 623–634. https://doi.org/10.1521/soco.2009.27.4.623

Kinzler, K. D., & Spelke, E. S. (2011). Do infants show social preferences for people differing in race? *Cognition*, *119*(1), 1–9. https://doi.org/10.1016/j.cognition.2010.10.019

Klauer, K. C., & Mierke, J. (2005). Task-Set Inertia, Attitude Accessibility, and Compatibility-Order Effects: New Evidence for a Task-Set Switching Account of the Implicit Association Test Effect. *Personality and Social Psychology Bulletin*, *31*(2), 208–217. https://doi.org/10.1177/0146167204271416

Kowatch, K., Whalen, J. M., & Pexman, P. M. (2013). Irony Comprehension in Action: A New Test of Processing for Verbal Irony. *Discourse Processes*, *50*(5), 301–315. https://doi.org/10.1080/0163853X.2013.799934

Kreuz, R. J. (2018). The use of verbal irony: Cues and constraints. In *Metaphor: Implications and applications* (pp. 23–38). Psychology Press.

Kreuz, R. J., & Link, K. E. (2002). Asymmetries in the use of verbal irony. *Journal of Language and Social Psychology*, *21*(2), 127–143. https://doi.org/10.1177/02627X02021002002

Kroczek, L. O. H., & Gunter, T. C. (2021). The time course of speaker-specific language processing. *Cortex*. https://doi.org/10.1016/j.cortex.2021.04.017

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, *47*(4), 2025–2047. https://doi.org/10.1007/s11135-011-9640-9

Kuperberg, G. R., Paczynski, M., & Ditman, T. (2011). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience*, *23*(5), 1230–1246. https://doi.org/10.1162/jocn.2010.21452

Kurtzman, H. S., & MacDonald, M. C. (1993). Resolution of quantifier scope ambiguities. *Cognition*, *48*(3), 243–279. https://doi.org/10.1016/0010-0277(93)90042-T

Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B., & Jensen, S. P. (2020). *lmerTest: Tests in Linear Mixed Effects Models* (Version 3.1-3) [Computer software].

Lai, V. T., Curran, T., & Menn, L. (2009). Comprehending conventional and novel metaphors: An ERP study. *Brain Research*, *1284*, 145–155. https://doi.org/10.1016/j.brainres.2009.05.088

Lakoff, G. (2002). *Moral politics: How liberals and conservatives think*. University of Chicago Press. https://search.library.wisc.edu/catalog/999926904702121

Lakoff, R. T. (1990). *Talking power: The politics of language*. Basic Books.

Lambert, W. E. (1967). A Social Psychology of Bilingualism. *Journal of Social Issues*, *23*(2), 91–109. https://doi.org/10.1111/j.1540-4560.1967.tb00578.x

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

Lane, H. (1963). Foreign accent and speech distortion. *The Journal of the Acoustical Society of America*, *35*(4).

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and Using the Implicit Association Test: IV: What We Know (So Far) about the Method. In *Implicit measures of attitudes.* (pp. 59–102). The Guilford Press.

Lavine, H., Lodge, M., Polichak, J., & Taber, C. (2002). Explicating the black box through experimentation: Studies of authoritarianism and threat. *Political Analysis*, *10*(04), 343–361. https://doi.org/10.1093/pan/10.4.343

Lee, R. (2015). Implicit Associations with Welsh in Two Educational Contexts. *York Papers in Linguistics*, *2*(14), 81–105.

Lenth, R. V., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (Version 1.8.9) [Computer software].

Lev-Ari, S. (2015). Comprehending non-native speakers: Theory and evidence for adjustment in manner of processing. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01546

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096. https://doi.org/10.1016/j.jesp.2010.05.025

Lev-Ari, S., & Keysar, B. (2012). Less-Detailed Representation of Non-Native Language: Why Non-Native Speakers' Stories Seem More Vague. *Discourse Processes*, *49*(7), 523–538. https://doi.org/10.1080/0163853X.2012.698493

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT press.

Li, B. B., & Huang, H.-W. (2024). Cognition and beyond: Intersections of personality traits and language. In *Psychology of Learning and Motivation* (Vol. 80, pp. 105–148). Elsevier. https://doi.org/10.1016/bs.plm.2024.03.003

Li, S., & Gu, W. (2021). *Prosodic Profiles of the Mandarin Speech Conveying Ironic Compliment*. 5. https://doi.org/10.1109/ISCSLP49672.2021.9362092

Li, S., Gu, W., Liu, L., & Tang, P. (2020). The Role of Voice Quality in Mandarin Sarcastic Speech: An Acoustic and Electroglottographic Study. *Journal of Speech, Language, and Hearing Research*, *63*(8), 2578–2588. https://doi.org/10.1044/2020_JSLHR-19-00166

Li, S., Jiang, X., Yu, H., & Zhou, X. (2014). Cognitive empathy modulates the processing of pragmatic constraints during sentence comprehension. *Social Cognitive and Affective Neuroscience*, *9*(8), 1166–1174. https://doi.org/10.1093/scan/nst091

Lindemann, S. (2005). Who speaks “broken English”? US undergraduates’ perceptions of non-native English 1. *International Journal of Applied Linguistics*, *15*(2), 187–212.

Lippi-Green, R. (2012). *English with an Accent: Language, Ideology and Discrimination in the United States* (2 (Online-Ausg.)). Taylor and Francis.

Llurda, E. (2009). The decline and fall of the native speaker teacher. *Language Teaching and Learning*, 37–53.

Locke, K. D. (2003). Status and solidarity in social comparison: Agentic and communal values and vertical and horizontal directions. *Journal of Personality and Social Psychology*, *84*(3), 619–631. https://doi.org/10.1037/0022-3514.84.3.619

Lorenzoni, A., Pagliarini, E., Vespignani, F., & Navarrete, E. (2022). Pragmatic and knowledge range lenience toward foreigners. *Acta Psychologica*, *226*, 103572. https://doi.org/10.1016/j.actpsy.2022.103572

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social Influence Effects on Automatic Racial Prejudice. *Journal of Personality and Social Psychology*, *81*(5), 842–855.

Luccasen, R. A., & Thomas, M. K. (2014). Monetary incentives versus class credit: Evidence from a large classroom trust experiment. *Economics Letters*, *123*(2), 232–235. https://doi.org/10.1016/j.econlet.2014.02.012

Lüdecke, D., Bartel, A., Schwemmer, C., Powell, C., Djalovski, A., & Titz, J. (2023). *sjPlot: Data Visualization for Statistics in Social Science* (Version 2.8.15) [Computer software]. https://cran.r-project.org/web/packages/sjPlot/index.html

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Macagno, F., & Bigi, S. (2018). Types of Dialogue and Pragmatic Ambiguity. In S. Oswald, T. Herman, & J. Jacquin (Eds.), *Argumentation and Language—Linguistic, Cognitive and Discursive Explorations* (Vol. 32, pp. 191–218). Springer International Publishing. https://doi.org/10.1007/978-3-319-73972-4_9

MacNair, L. (2021). Implicit bias and perception of accent. *Working Papers of the Linguistics Circle*, *31*(1).

Macrae, C. N., & Bodenhausen, G. V. (2000). Social Cognition: Thinking Categorically about Others. *Annual Review of Psychology*, *51*(1), 93–120. https://doi.org/10.1146/annurev.psych.51.1.93

Mai, R., & Hoffmann, S. (2014). Accents in Business Communication: An integrative model and propositions for future research. *Journal of Consumer Psychology*, *24*(1), 137–158. https://doi.org/10.1016/j.jcps.2013.09.004

Major, R. C. (2007). Identifying a foreign accent in an unfamiliar language. *Studies in Second Language Acquisition*, *29*(04). https://doi.org/10.1017/S0272263107070428

Mange, J., Lepastourel, N., & Georget, P. (2009). Is Your Language a Social Clue? Lexical Markers and Social Identity. *Journal of Language and Social Psychology*, *28*(4), 364–380. https://doi.org/10.1177/0261927X09341956

Markman, A. B., Taylor, E., & Gentner, D. (2007). Auditory presentation leads to better analogical retrieval than written presentation. *Psychonomic Bulletin & Review*, *14*(6), 1101–1106. https://doi.org/10.3758/BF03193098

Marrville, C. (2017). *Gender and dominance in action: World view and emotional affect in language processing and use*.

Martin, C. D., Garcia, X., Breton, A., Thierry, G., & Costa, A. (2014). From literal meaning to veracity in two hundred milliseconds. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00040

Martin, I., & McDonald, S. (2003). Weak coherence, no theory of mind, or executive dysfunction? Solving the puzzle of pragmatic language disorders. *Brain and Language*, *85*(3), 451–466. https://doi.org/10.1016/S0093-934X(03)00070-1

Mauchand, M., Vergis, N., & Pell, M. D. (2018). *Ironic tones of voices*. 5.

Mauchand, M., Vergis, N., & Pell, M. D. (2019). Irony, Prosody, and Social Impressions of Affective Stance. *Discourse Processes*, 1–17. https://doi.org/10.1080/0163853x.2019.1581588

Mayn, A., & Demberg, V. (2022). Individual Differences in a Pragmatic Reference Game. *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.

McCulloch, G. (2019). *Because internet: Understanding the new rules of language*. Riverhead Books.

McGlone, M. S., & Tofighbakhsh, J. (2000). Birds of a Feather Flock Conjointly (?): Rhyme as Reason in Aphorisms. *Psychological Science*, *11*(5), 424–428. https://doi.org/10.1111/1467-9280.00282

McLaughlin, D. J., Baese-Berk, M. M., Bent, T., Borrie, S. A., & Van Engen, K. J. (2018). Coping with adversity: Individual differences in the perception of noisy and accented speech. *Attention, Perception, & Psychophysics*, *80*(6), 1559–1570. https://doi.org/10.3758/s13414-018-1537-4

Medgyes, P. (1992). Native or non-native: Who's worth more? *ELT Journal*, *46*(4), 340–349. https://doi.org/10.1093/elt/46.4.340

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, *90*(5), 862–877. https://doi.org/10.1037/0022-3514.90.5.862

Meo, A. D., Vitale, M., Pettorino, M., & Martin, P. (2011). Acoustic-perceptual credibility correlates of news reading by native and Chinese speakers of Italian. *Hong Kong*.

Mey, J. L. (2001). *Pragmatics: An introduction*.

Mey, J. L. (2003). Context and (dis)ambiguity: A pragmatic view. *Journal of Pragmatics*.

Mikhailova, P. (2020). *Graphic markers of irony and sarcasm in written texts*. 346–356.

Milosky, L. M., & Wrobleski, C. A. (1994). *The prosody of irony*. International Society for Humor Studies Conference, Ithaca, NY.

Mischel, W. (2013). *Personality and assessment*. Psychology Press.

Molenberghs, P., & Louis, W. R. (2018). Insights From fMRI Studies Into Ingroup Bias. *Frontiers in Psychology*, *9*, 1868. https://doi.org/10.3389/fpsyg.2018.01868

Munro, M. J. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition*, *20*(2), 139–154. https://doi.org/10.1017/S0272263198002022

Munro, M. J., & Derwing, T. M. (1995). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning*, *45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Münster, K., & Knoeferle, P. (2018). Extending Situated Language Comprehension (Accounts) with Speaker and Comprehender Characteristics: Toward Socially Situated Interpretation. *Frontiers in Psychology*, *8*, 2267. https://doi.org/10.3389/fpsyg.2017.02267

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language Discrimination by English-Learning 5-Month-Olds: Effects of Rhythm and Familiarity. *Journal of Memory and Language*, *43*(1), 1–19. https://doi.org/10.1006/jmla.2000.2698

Neuberg, S. L., & Schaller, M. (2008). Intergroup prejudices and intergroup conflicts. *Found. Evolut. Psychol*, *401*.

Niedzielski, N. (1999). The Effect of Social Information on the Perception of Sociolinguistic Variables. *Journal of Language and Social Psychology*, *18*(1), 62–85. https://doi.org/10.1177/0261927X99018001005

Niemi, L., Hartshorne, J., Gerstenberg, T., Stanley, M., & Young, L. (2020). Moral Values Reveal the Causality Implicit in Verb Meaning. *Cognitive Science*, *44*(6), e12838. https://doi.org/10.1111/cogs.12838

Nisbett, R. E., & Wilson, T. D. (1977). The Halo Effect: Evidence for Unconscious Alteration of Judgments. *Journal of Personality and Social Psychology*, *35*(4), 250–256. https://doi.org/doi:10.1037/0022-3514.35.4.250

Obert, A., Gierski, F., & Caillies, S. (2018). He catapulted his words from the dais: An ERP investigation of novel verbal metaphors. *Journal of Neurolinguistics*, *47*, 59–70. https://doi.org/10.1016/j.jneuroling.2018.02.008

Ogden, D. (2020). *Processing Fluency, Perceptual Adaptation, and Language Attitudes: Does Adaptation Improve Comprehension Ease and Attitudes Toward Speakers with Non-Native Accents?*

Ogunnaike, O., Dunham, Y., & Banaji, M. R. (2010). The language of implicit preferences. *Journal of Experimental Social Psychology*, *46*(6), 999–1003. https://doi.org/10.1016/j.jesp.2010.07.006

Olkoniemi, H., Ranta, H., & Kaakinen, J. K. (2016). Individual differences in the processing of written sarcasm and metaphor: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(3), 433–450. https://doi.org/10.1037/xlm0000176

Olson, M. A., & Fazio, R. H. (2004). Reducing the Influence of Extrapersonal Associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology*, *86*(5), 653–667. https://doi.org/10.1037/0022-3514.86.5.653

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237–241. https://doi.org/10.1016/j.tics.2008.02.014

Oppenheimer, D. M., & Frank, M. C. (2008). A rose in any other font would not smell as sweet: Effects of perceptual fluency on categorization. *Cognition*, *106*(3), 1178–1194. https://doi.org/10.1016/j.cognition.2007.05.010

O'Rourke, B., & Pujolar, J. (2013). *From native speakers to "new speakers"- Problematizing nativeness in language revitalization contexts*.

Osborne, D., Costello, T. H., Duckitt, J., & Sibley, C. G. (2023). The psychological causes and societal consequences of authoritarianism. *Nature Reviews Psychology*, *2*(4), 220–232. https://doi.org/10.1038/s44159-023-00161-4

Osterhout, L., Bersick, M., & Mclaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, *25*(3), 273–285. https://doi.org/10.3758/BF03211283

Oxley, D. R., Smith, K. B., Alford, J. R., Hibbing, M. V., Miller, J. L., Scalora, M., Hatemi, P. K., & Hibbing, J. R. (2008). Political Attitudes Vary with Physiological Traits. *Science*, *321*(5896), 1667–1670. https://doi.org/10.1126/science.1157627

Padilla, G. X. (2011). ¿Existen rasgos prosódicos objetivos en los enunciados irónicos? *Revista Oralia*, *14*, 203–227.

Paladino, M. P., & Mazzurega, M. (2020). One of Us: On the Role of Accent and Race in Real-Time In-Group Categorization. *Journal of Language and Social Psychology*, *39*(1), 22–39. https://doi.org/10.1177/0261927X19884090

Paulmann, S., Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00345

Pehrson, S., Vignoles, V. L., & Brown, R. (2009). National Identification And Anti-Immigrant Prejudice: Individual And Contextual Effects Of National Definitions. *Social Psychology Quarterly*, *72*(1), 24–38. https://doi.org/10.1177/019027250907200104

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pélissier, M., & Ferragne, E. (2022). The N400 reveals implicit accent-induced prejudice. *Speech Communication*, *137*, 114–126. https://doi.org/10.1016/j.specom.2021.10.004

Petropoulos Petalas, D., Schumacher, G., & Scholte, S. H. (2024). Is political ideology correlated with brain structure? A preregistered replication. *iScience*, 110532. https://doi.org/10.1016/j.isci.2024.110532

Pexman, P. M. (2008). It's fascinating research: The cognition of verbal irony. *Current Directions in Psychological Science*, *17*(4), 286–290. https://doi.org/10.1111/j.1467-8721.2008.00591.x

Pexman, P. M., & Zvaigzne, M. T. (2004). Does irony go better with friends? *Metaphor and Symbol*, *19*(2), 143–163. https://doi.org/10.1207/s15327868ms1902_3

Phillips, M. L., Young, A. W., Scott, S. K., Calder, A. J., Andrew, C., Giampietro, V., Williams, S. C. R., Bullmore, E. T., Brammer, M., & Gray, J. A. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1408), 1809–1817. https://doi.org/10.1098/rspb.1998.0506

Pietraszewski, D., & Schwartz, A. (2014). Evidence that accent is a dedicated dimension of social categorization, not a byproduct of coalitional categorization. *Evolution and Human Behavior*, *35*(1), 51–57. https://doi.org/10.1016/j.evolhumbehav.2013.09.005

Piller, I. (2002). Passing for a native speaker: Identity and success in second language learning. *Journal of Sociolinguistics*, *6*(2), 179–208. https://doi.org/10.1111/1467-9481.00184

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2021). *nlme: Linear and Nonlinear Mixed Effects Models* (Version R package version 3.1-149) [Computer software]. https://CRAN.R-project.org/package=nlme

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*(2), 191–215. https://doi.org/10.1006/jpho.2001.0134

Podlipský, V. J., Šimáčková, Š., & Petráž, D. (2016). Is there an interlanguage speech credibility benefit? *Topics in Linguistics*, *17*(1), 30–44. https://doi.org/10.1515/topling-2016-0003

Porretta, V., Buchanan, L., & Järvikivi, J. (2020). When processing costs impact predictive processing: The case of foreign-accented speech and accent experience. *Attention, Perception, & Psychophysics*, *82*(4), 1558–1565. https://doi.org/10.3758/s13414-019-01946-7

Porretta, V., Tucker, B. V., & Järvikivi, J. (2016). The influence of gradient foreign accentedness and listener experience on word recognition. *Journal of Phonetics*, *58*, 1–21. https://doi.org/10.1016/j.wocn.2016.05.006

Puhacheuskaya, V., & Järvikivi, J. (2021). *Final dataset for "I was being sarcastic!: The effect of foreign accent and political ideology on irony (mis)understanding."* [Computer software]. osf.io/2dhbt

Puhacheuskaya, V., & Järvikivi, J. (2022). I was being sarcastic!: The effect of foreign accent and political ideology on irony (mis)understanding. *Acta Psychologica*, *222*, 103479. https://doi.org/10.1016/j.actpsy.2021.103479

Pycha, A., & Zellou, G. (2024). The influence of accent and device usage on perceived credibility during interactions with voice-AI assistants. *Frontiers in Computer Science*, *6*, 1411414. https://doi.org/10.3389/fcomp.2024.1411414

R Core Team. (2020). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rakić, T., Steffens, M. C., & Mummendey, A. (2011). Blinded by the accent! The minor role of looks in ethnic categorization. *Journal of Personality and Social Psychology*, *100*(1), 16–29. https://doi.org/10.1037/a0021522

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, *8*(3), 338–342.

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of Perceptual Fluency on Affective Judgments. *Psychological Science*, *9*(1), 45–48. https://doi.org/10.1111/1467-9280.00008

Regel, S., Coulson, S., & Gunter, T. C. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain Research*, *1311*, 121–135. https://doi.org/10.1016/j.brainres.2009.10.077

Regel, S., Gunter, T. C., & Friederici, A. D. (2011). Isn't it ironic? An electrophysiological exploration of figurative language processing. *Journal of Cognitive Neuroscience*, *23*(2), 277–293. https://doi.org/10.1162/jocn.2010.21411

Regel, S., Meyer, L., & Gunter, T. C. (2014). Distinguishing neurocognitive processes reflected by P600 effects: Evidence from ERPs and neural oscillations. *PLoS ONE*, *9*(5), e96840. https://doi.org/10.1371/journal.pone.0096840

Reichl, A. J. (1997). Ingroup favouritism and outgroup favouritism in low status minimal groups: Differential responses to statusrelated and statusunrelated measures. *European Journal of Social Psychology*, *27*(6), 617–633. https://doi.org/10.1002/(sici)1099-0992(199711/12)27:6<617::aid-ejsp829>3.0.co;2-t

Richetin, J., Costantini, G., Perugini, M., & Schönbrodt, F. (2015). Should We Stop Looking for a Better Scoring Algorithm for Handling Implicit Association Test Data? Test of the Role of Errors, Extreme Latencies Treatment, Scoring Formula, and Practice Trials on

Reliability and Validity. *PLOS ONE*, *10*(6), e0129601. https://doi.org/10.1371/journal.pone.0129601

Rockwell, P. (2000). Lower, slower, louder: Vocal cues of sarcasm. *Journal of Psycholinguistic Research*, *29*(5), 483–495.

Rockwell, P. (2007). Vocal Features of Conversational Sarcasm: A Comparison of Methods. *J Psycholinguist Res*, 9.

Rockwell, P., & Theriot, E. M. (2001). Culture, Gender, and Gender Mix in Encoders of Sarcasm: A Self- Assessment Analysis. *Communication Research Reports*, *18*(1), 44–52. https://doi.org/10.1080/08824090109384781

Roessel, J., Schoel, C., & Stahlberg, D. (2018). What's in an accent? General spontaneous biases against nonnative accents: An investigation with conceptual and auditory IATs: General biases against nonnative accents. *European Journal of Social Psychology*, *48*(4), 535–550. https://doi.org/10.1002/ejsp.2339

Romero-Rivas, C., Martin, C. D., & Costa, A. (2015). Processing changes when listening to foreign-accented speech. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00167

Romero-Rivas, C., Martin, C. D., & Costa, A. (2016). Foreign-accented speech modulates linguistic anticipatory processes. *Neuropsychologia*, *85*, 245–255. https://doi.org/10.1016/j.neuropsychologia.2016.03.022

Romero-Rivas, C., Morgan, C., & Collier, T. (2022). Accentism on Trial: Categorization/Stereotyping and Implicit Biases Predict Harsher Sentences for Foreign-Accented Defendants. *Journal of Language and Social Psychology*, *41*(2), 191–208. https://doi.org/10.1177/0261927X211022785

Rule, N. O., Adams, R. B., Ambady, N., & Freeman, J. B. (2012). Perceptions of Dominance following Glimpses of Faces and Bodies. *Perception*, *41*(6), 687–706. https://doi.org/10.1068/p7023

Ryan, E. B. (1983). Social Psychological Mechanisms Underlying Native Speaker Evaluations of Non-Native Speech. *Studies in Second Language Acquisition*, *5*(2), 148–159. https://doi.org/10.1017/S0272263100004824

Ryan, E. B., & Carranza, M. A. (1975). Evaluative reactions of adolescents toward speakers of standard English and Mexican American accented English. *Journal of Personality and Social Psychology*, *31*(5), 855–863. https://doi.org/10.1037/h0076704

Ryan, E. B., & Sebastian, R. J. (1980). The effects of speech style and social class background on social judgements of speakers. *British Journal of Social and Clinical Psychology*, *19*(3), 229–233.

Satpute, A. B., & Lieberman, M. D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research*, *1079*(1), 86–97. https://doi.org/10.1016/j.brainres.2006.01.005

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, *107*(6), 2408–2412. https://doi.org/10.1073/pnas.0908239106

Schirmer, A., Escoffier, N., Zysset, S., Koester, D., Striano, T., & Friederici, A. D. (2008). When vocal processing gets emotional: On the role of social orientation in relevance detection by the human amygdala. *NeuroImage*, *40*(3), 1402–1410. https://doi.org/10.1016/j.neuroimage.2008.01.018

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, *14*(2), 228–233. https://doi.org/10.1016/S0926-6410(02)00108-8

Schwarz, N. (2004). Metacognitive Experiences in Consumer Judgment and Decision Making. *Journal of Consumer Psychology*, *14*(4), 332–348. https://doi.org/10.1207/s15327663jcp1404_2

Scott, S. K. (2019). From speech and talkers to the social world: The neural processing of human spoken language. *Science*, *366*(6461), 58–62.

Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics*, *20*, 213–223. https://doi.org/10.1017/S0267190500200135

Searle, J. R. (1975). *A taxonomy of illocutionary acts*.

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge University Press.

Shamay-Tsoory, S. G., Tomer, R., Berger, B. D., Goldsher, D., & Aharon-Peretz, J. (2005). Impaired "affective Theory of Mind" is associated with right ventromedial prefrontal damage. *Cognitive and Behavioral Neurology*, *18*(1), 55–67. https://doi.org/10.1097/01.wnn.0000152228.90129.99

Sibley, C. G., Osborne, D., & Duckitt, J. (2012). Personality and political orientation: Meta-analysis and test of a Threat-Constraint Model. *Journal of Research in Personality*, *46*(6), 664–677. https://doi.org/10.1016/j.jrp.2012.08.002

Skitka, L. J., & Sargis, E. G. (2006). The Internet as Psychological Laboratory. *Annual Review of Psychology*, *57*(1), 529–555. https://doi.org/10.1146/annurev.psych.57.102904.190048

Slugoski, B. R., & Turnbull, W. (1988). Cruel to be kind and kind to be cruel: Sarcasm, banter and social relations. *Journal of Language and Social Psychology*, *7*(2), 101–121. https://doi.org/10.1177/0261927X8800700202

Soto, C. J. (2019). How Replicable Are Links Between Personality Traits and Consequential Life Outcomes? The Life Outcomes of Personality Replication Project. *Psychological Science*, *30*(5), 711–727. https://doi.org/10.1177/0956797619831612

Soto, C. J. (2021). Do Links Between Personality and Life Outcomes Generalize? Testing the Robustness of Trait–Outcome Associations Across Gender, Age, Ethnicity, and Analytic Approaches. *Social Psychological and Personality Science*, *12*(1), 118–130. https://doi.org/10.1177/1948550619900572

Souza, A. L. (2012). *Modality-specific effects of processing fluency on cognitive judgments*. University of Texas.

Souza, A. L., & Markman, A. B. (2013). Foreign accent does not influence cognitive judgments. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *35*.

Spence, J. L., Hornsey, M. J., Stephenson, E. M., & Imuta, K. (2024). Is Your Accent Right for the Job? A Meta-Analysis on Accent Bias in Hiring Decisions. *Personality and Social Psychology Bulletin*, *50*(3), 371–386. https://doi.org/10.1177/01461672221130595

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J.-B., & Noveck, I. A. (2012). Neural evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage*, *63*(1), 25–39. https://doi.org/10.1016/j.neuroimage.2012.06.046

Stamenković, D., Ichien, N., & Holyoak, K. J. (2019). Metaphor comprehension: An individual-differences approach. *Journal of Memory and Language*, *105*, 108–118. https://doi.org/10.1016/j.jml.2018.12.003

Statistics Canada. (2017). *Edmonton, CY [Census subdivision], Alberta and Division No. 11, CDR [Census division], Alberta (table). Census Profile. 2016 Census. Statistics Canada Catalogue no. 98-316-X2016001. Released November 29, 2017.* Ottawa. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E

Steen, G. J. (2023). Thinking by metaphor, fast and slow: Deliberate Metaphor Theory offers a new model for metaphor and its comprehension. *Frontiers in Psychology*, *14*, 1242888. https://doi.org/10.3389/fpsyg.2023.1242888

Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A., Krennmayr, T., & Pasma, T. (2010). *A method for linguistic metaphor identification: From MIP to MIPVU* (Vol. 14). John Benjamins Publishing.

Stern, H. H. (2003). *Fundamental concepts of language teaching* (12. impr). Oxford Univ. Press.

Stieger, S., & Reips, U.-D. (2016). A limitation of the Cognitive Reflection Test: Familiarity. *PeerJ*, *4*, e2395. https://doi.org/10.7717/peerj.2395

Stocker, L. (2017). The Impact of Foreign Accent on Credibility: An Analysis of Cognitive Statement Ratings in a Swiss Context. *Journal of Psycholinguistic Research*, *46*(3), 617–628. https://doi.org/10.1007/s10936-016-9455-x

Stringaris, A. K., Medford, N. C., Giampietro, V., Brammer, M. J., & David, A. S. (2007). Deriving meaning: Distinct neural mechanisms for metaphoric, literal, and non-meaningful

sentences. *Brain and Language*, *100*(2), 150–162. https://doi.org/10.1016/j.bandl.2005.08.001

Sztencel, M. (2018). Semantics, Pragmatics and Meaning Revisited. *Cham: Springer International Publishing. Doi: Https://Doi. Org/10.1007/978-3-319-69116-9*.

Tahta, S., Wood, M., & Loewenthal, K. (1981). Foreign Accents: Factors Relating to Transfer of Accent from the First Language to a Second Language. *Language and Speech*, *24*(3), 265–272. https://doi.org/10.1177/002383098102400306

Taifel, H. (1981). Human groups and social categories. *Studies in Social Psychology*.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, *223*(5), 96–103.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178. https://doi.org/10.1002/ejsp.2420010202

Tajfel, H., & Turner, J. (1986). *The social identity theory of intergroup behavior*. *5*, 7–24.

Tartter, V. C., Gomes, H., Dubrovsky, B., Molholm, S., & Stewart, R. V. (2002). Novel Metaphors Appear Anomalous at Least Momentarily: Evidence from N400. *Brain and Language*, *80*(3), 488–509. https://doi.org/10.1006/brln.2001.2610

Tesink, C. M. J. Y., Petersson, K. M., van Berkum, J. J. A., van den Brink, D., Buitelaar, J. K., & Hagoort, P. (2009). Unification of speaker and meaning in language comprehension: An fMRI study. *Journal of Cognitive Neuroscience*, *21*(11), 2085–2099. https://doi.org/10.1162/jocn.2008.21161

Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic Segment Durations in Prosodic Research: A Practical Guide. In S. Sudhoff, D. Lenertova, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, & J. Schließer (Eds.), *Methods in Empirical Prosody Research* (pp. 1–28). DE GRUYTER. https://doi.org/10.1515/9783110914641.1

Unkelbach, C. (2007). Reversing the truth effect: Learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 219–230. https://doi.org/10.1037/0278-7393.33.1.219

Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2008). The Neural Substrates of In-Group Bias: A Functional Magnetic Resonance Imaging Investigation. *Psychological Science*, *19*(11), 1131–1139. https://doi.org/10.1111/j.1467-9280.2008.02214.x

Van Bavel, J. J., Packer, D. J., & Cunningham, W. A. (2011). Modulation of the Fusiform Face Area following Minimal Exposure to Motivationally Relevant Faces: Evidence of In-group Enhancement (Not Out-group Disregard). *Journal of Cognitive Neuroscience*, *23*(11), 3343–3354. https://doi.org/10.1162/jocn_a_00016

Van Berkum, J. J. A. (2009). The neuropragmatics of "simple" utterance comprehension: An ERP review. In *Semantics and pragmatics: From experiment to theory* (pp. 276–316). Basingstoke: Palgrave Macmillan.

van Berkum, J. J. A. (2018). *Language comprehension, emotion, and sociality* (S.-A. Rueschemeyer & M. G. Gaskell, Eds.; 2nd ed.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198786825.013.28

Van Berkum, J. J. A., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009a). Right or wrong?: The brain's fast response to morally objectionable statements. *Psychological Science*, *20*(9), 1092–1099. https://doi.org/10.1111/j.1467-9280.2009.02411.x

Van Berkum, J. J. A., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009b). Right or Wrong?: The Brain's Fast Response to Morally Objectionable Statements. *Psychological Science*, *20*(9), 1092–1099. https://doi.org/10.1111/j.1467-9280.2009.02411.x

Van Berkum, J. J. A., Van Den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The Neural Integration of Speaker and Message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591. https://doi.org/10.1162/jocn.2008.20054

Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The Neural Integration of Speaker and Message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591. https://doi.org/10.1162/jocn.2008.20054

van den Brink, D., Van Berkum, J. J. A., Bastiaansen, M. C. M., Tesink, C. M. J. Y., Kos, M., Buitelaar, J. K., & Hagoort, P. (2012a). Empathy matters: ERP evidence for inter-

individual differences in social language processing. *Social Cognitive and Affective Neuroscience*, *7*(2), 173–183. https://doi.org/10.1093/scan/nsq094

van den Brink, D., Van Berkum, J. J. A., Bastiaansen, M. C. M., Tesink, C. M. J. Y., Kos, M., Buitelaar, J. K., & Hagoort, P. (2012b). Empathy matters: ERP evidence for inter-individual differences in social language processing. *Social Cognitive and Affective Neuroscience*, *7*(2), 173–183. https://doi.org/10.1093/scan/nsq094

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00577

Van Maastricht, L., Krahmer, E., & Swerts, M. (2016). Prominence Patterns in a Second Language: Intonational Transfer From Dutch to Spanish and Vice Versa. *Language Learning*, *66*(1), 124–158. https://doi.org/10.1111/lang.12141

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2020). *Itsadug: Interpreting time series and autocorrelated data using GAMMs. R package version 2.4* [Computer software]. https://cran.r-project.org/web/packages/itsadug/index.html

Vissers, C. Th. W. M., Chwilla, U. G., Egger, J. I. M., & Chwilla, D. J. (2013). The interplay between mood and language comprehension: Evidence from P600 to semantic reversal anomalies. *Neuropsychologia*, *51*(6), 1027–1039. https://doi.org/10.1016/j.neuropsychologia.2013.02.007

Voeten, C. C. (2021). *buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression* (Version R package version 2.2) [Computer software]. https://cran.r-project.org/web/packages/buildmer/index.html

Weatherholtz, K., Campbell-Kibler, K., & Jaeger, T. F. (2014). Socially-mediated syntactic alignment. *Language Variation and Change*, *26*(3), 387–420. https://doi.org/10.1017/S0954394514000155

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, *67*(6), 1049–1062. https://doi.org/10.1037/0022-3514.67.6.1049

Werkmann Horvat, A., Bolognesi, M., Littlemore, J., & Barnden, J. (2022). Comprehension of different types of novel metaphors in monolinguals and multilinguals. *Language and Cognition*, *14*(3), 401–436. https://doi.org/10.1017/langcog.2022.8

Wesołek, S., Gulgowski, P., Błaszczak, J., & Żygis, M. (2023). *Illusions of Ungrammaticality in the Perception of Non-Native Accented Speech*. 20th International Congress of Phonetic Sciences (ICPhS), Prague.

Wester, F., Gilbers, D., & Lowie, W. (2007). Substitution of dental fricatives in English by Dutch L2 speakers. *Language Sciences*, *29*(2–3), 477–491. https://doi.org/10.1016/j.langsci.2006.12.029

Wetzel, M. (2021). Do non-native and unfamiliar accents sound less credible? An examination of the processing. *Journal of Articles in Support of the Null Hypothesis*, *17*(2).

Weyant, J. M. (2007). Perspective Taking as a Means of Reducing Negative Stereotyping of Individuals Who Speak English as a Second Language. *Journal of Applied Social Psychology*, *37*(4), 703–716. https://doi.org/10.1111/j.1559-1816.2007.00181.x

Whalen, J. M., Pexman, P. M., & Gill, A. J. (2009). "Should Be Fun—Not!": Incidence and Marking of Nonliteral Language in E-Mail. *Journal of Language and Social Psychology*, *28*(3), 263–280. https://doi.org/10.1177/0261927X09335253

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., & Dunnington, D. (2023). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics* (Version 3.4.4) [Computer software]. https://cran.r-project.org/web/packages/ggplot2/index.html

Wilson, G. D., & Patterson, J. R. (1968). A new measure of conservatism. *British Journal of Social and Clinical Psychology*, *7*(4), 264–269. https://doi.org/10.1111/j.2044-8260.1968.tb00568.x

Winkielman, P., & Cacioppo, J. T. (2001). Mind at ease puts a smile on the face: Psychophysiological evidence that processing facilitation elicits positive affect. *Journal of Personality and Social Psychology*, *81*(6), 989.

Winner, E. (1993). Metaphor and irony: Two levels of understanding. *Metaphor and Thought/Ortony*.

Winner, E., Levy, J., Kaplan, J., & Rosenblatt, E. (1988). Children's understanding of nonliteral language. *Journal of Aesthetic Education*, *22*(1), 51. https://doi.org/10.2307/3332964

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, *75*(3), 537–556. https://doi.org/10.3758/s13414-012-0404-y

Wlotko, E. W., Federmeier, K. D., & Kutas, M. (2012). To predict or not to predict: Age-related differences in the use of sentential context. *Psychology and Aging*, *27*(4), 975–988. https://doi.org/10.1037/a0029206

Wood, N. S. (2017). *Generalized additive models: An introduction with R* (2 edition). Chapman and Hall/CRC. https://cran.r-project.org/web/packages/mgcv/citation.html

Wu, C. H. (1995). Language, interaction and social cognition. In *Applied Psycholinguistics* (Vol. 16, pp. 349–353). Sage. https://www.cambridge.org/core/product/identifier/S014271640006598X/type/journal_article

Yi, H.-G., Phelps, J. E. B., Smiljanic, R., & Chandrasekaran, B. (2013). Reduced efficiency of audiovisual integration for nonnative speech. *The Journal of the Acoustical Society of America*, *134*(5), EL387–EL393. https://doi.org/10.1121/1.4822320

Yi, H.-G., Smiljanic, R., & Chandrasekaran, B. (2014). The neural processing of foreign-accented speech and its relationship to listener bias. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00768

Yzerbyt, V. Y., Leyens, J., & Bellour, F. (1995). The ingroup overexclusion effect: Identity concerns in decisions about group membership. *European Journal of Social Psychology*, *25*(1), 1–16. https://doi.org/10.1002/ejsp.2420250102

Zhou, P., & Christianson, K. (2016a). Auditory perceptual simulation: Simulating speech rates or accents? *Acta Psychologica*, *168*, 85–90. https://doi.org/10.1016/j.actpsy.2016.04.005

Zhou, P., & Christianson, K. (2016b). I "hear" what you're "saying": Auditory perceptual simulation, reading speed, and reading comprehension. *Quarterly Journal of Experimental Psychology*, *69*(5), 972–995. https://doi.org/10.1080/17470218.2015.1018282

Zhou, P., Garnsey, S., & Christianson, K. (2019). Is imagining a voice like listening to it? Evidence from ERPs. *Cognition*, *182*, 227–241. https://doi.org/10.1016/j.cognition.2018.10.014

Zink, C. F., Tong, Y., Chen, Q., Bassett, D. S., Stein, J. L., & Meyer-Lindenberg, A. (2008). Know Your Place: Neural Processing of Social Hierarchy in Humans. *Neuron*, *58*(2), 273–283. https://doi.org/10.1016/j.neuron.2008.01.025

Appendix A Stimuli

A.1 Final stimuli for Chapter 2

Ironic Praise / Literal Praise
Literal Criticism / Ironic Criticism

2.

I saw on Facebook that you had a cello concert in Vienna.

Yes, they called me for an encore three times, and the applause was so loud. I stayed there till midnight.

I always knew | you were | **hopeless/gifted** | at playing | the cello

How are your cello lessons going? Are you getting better?

It's torture. Last time the teacher kicked me out. She said she could give me a badge for being her worst student ever.

I always knew | you were | **hopeless/gifted** | at playing | the cello

4.

Have you decided what to do with the cash we found? I deposited my share.

Yeah, I donated half to the local food bank and used the other half to get my grandma the new rocking chair she's been dreaming of.

You are always | incredibly | **stingy/generous** | about giving | to others

We are all chipping in for a farewell gift for Ms. Walsh. Are you on board?

Please, I'm not a billionaire. Why should I waste money on dust-collecting souvenirs? Besides, I think I forgot my wallet.

You are always | incredibly | **stingy/generous**| about giving | to others

7.

So what is your background in biology? You make very strong claims.

I have two Master's degrees, one in Biotech and one in Computer Science, as well as a Ph.D. in Bioinformatics from Oxford. I'm now a postdoc at MIT.

You are really | profoundly | **ignorant/knowledgeable** | about this | subject

The economy is in deep crisis. We're heading for default.

I am no economist, but I never understood why they couldn't just print more money and repay all debts. That sounds pretty simple, doesn't it?

You are really | profoundly | **ignorant/knowledgeable**| about this | subject

8.

Is there any milk left in the fridge? I totally forgot it was my turn to shop.

There were huge discounts on dairy and fresh produce at the farmer's market yesterday so I stocked up the fridge for weeks.

As always, | you are so | **unreliable/reliable** | in small | matters

Did you get all the salad ingredients? I made the dressing.

I could not find any veggies in the nearest store and didn't feel like going anywhere else. I got some McDonald's instead.

As always, | you are so | **unreliable/reliable**| in small | matters

9.

Did your building managers get back to you about the low water pressure?

No, I've been waiting for a reply for a month already. I can't even rinse soap off my body, let alone shampoo. But I sort of got used to doing it at the gym or work.

I'm impressed | you are so | **impatient/patient** | about | this issue

Did they manage to fix the problem with the minibar in your room?

Don't rub it in. I sent them a dozen emails and even came down to yell at the receptionist. I swear if by the time I get in it is not solved I'm calling the manager.

I am impressed | you are so | **impatient/patient** | about | this issue

10.

They just announced the results of a grant competition. My proposal was rejected for the second year in a row.

Well, I got a thirty-thousand-dollar grant for my Ph.D. project on ground beetles.

Your supervisor | must be | **ashamed/proud** | of your work | right now

Are you enjoying your Ph.D. classes so far? You're in Economics, right?

I got caught plagiarizing my term paper on Econometrics. They gave me an F and put me on academic probation. I'm considering quitting.

Your supervisor | must be | **ashamed/proud**| of your work | right now

11.

How did the dinner with your wife's parents go?

I spent the whole day cooking salmon and making drinks. They finished everything, down to the last crumb, and asked for seconds.

You surely are | the most | **terrible/talented** | cook in this | part of town

How was your Saturday dinner with your wife?

I decided to try a new fish recipe. I probably messed up because she couldn't finish it and the next day got really bad food poisoning.

You surely are | the most | **terrible/talented**| cook in this | part of town

13.

I have no idea when to start working on our group project. I'm throwing a party this weekend.

Don't worry. I already did a full market analysis and came up with a valid predictive model. Just double-check it.

You truly | are the | **laziest/keenest** | student | among us

When are you planning to turn in your part of the group project? Everyone has already sent theirs.

Give me a break, will you? I'm binge-watching Dark on Netflix. It'll be done when it's done.

You are | truly the | **laziest/keenest** | student | among us

14.

Have you told her you got accepted to the top three schools in the U.S. with a full scholarship and tuition waiver?

Come on, there is nothing to talk about. It's not a big deal or anything.

You should try | being less | **arrogant/modest** | about | yourself

What were you arguing about with the boss?

You all don't respect me enough. I am literally the best thing that has ever happened to this business. No one else here has even a tenth of my skills.

You should try | being less | **arrogant/modest** | about | yourself

15.

I wish my parents pushed me more with sports at a young age. I could really have used that.

Oh, I had a passion for basketball in high school. I was team captain. Still have a whole cupboard of medals from tournaments. And you know, girls like athletes.

I see you were | a big-time | **loser/winner** | in high | school

Would you like to play football with us this weekend? It's gonna be fun!

I tried playing it in high school. I was so bad my mates didn't let me in their games. I missed every ball I could as goalie and never scored a single goal as forward.

I see you were | a big-time | **loser/winner** | in high | school

16.

Are these photos of your new house? I see you've redesigned almost everything.

Yes, I'm doing it all country style. I matched the colors of every single item for coordination. I've sent pictures to the journal of interior design and they are going to publish them.

How come | you are so | **tasteless/tasteful** | in home | decor?

Are those photos of your new apartment? How are you going to decorate it?

Oh, I don't care at all. Maybe I'll get a couple of second-hand sofas and put some old journal pictures on the walls.

How come | you are so | **tasteless/tasteful** | in home | decor?

17.

Did you manage to get into the museum yesterday? The line was so long.

The guard said I would be the last person they let in. There was an elderly gentleman behind me who got extremely upset so I let him in instead of me.

As always, | you are so | **inconsiderate/considerate** | to random | strangers

I didn't manage to get inside the Sistine Chapel yesterday. The line was ridiculous!

I pretended a guy in the front of the line was my friend and slipped inside before people realized what happened.

As always, | you are so | **inconsiderate/considerate** | to random | strangers

18.

I feel like a total failure. My marriage is falling apart, and my boss is going to lay me off.

You are a wonderful person with a bright future! It's just that sometimes things have to fall apart to make way for something better. Put on a smile.

Wow, you are | the most | **cheerless/cheerful** | person | I know

Summer is coming! I'm really looking forward to this vacation. What are your plans?

Getting the coronavirus, spending two wonderful weeks in the hospital, maybe even staying alive. The usual. Get ready for the same, by the way.

Wow, you are | the most | **cheerless/cheerful** | person | I know

19.

Did you check out that new park close to your place?

I go there regularly. There are so many adorable dogs and cats there! I always try to bring them some treats or toys to play with.

You are always| incredibly | **hostile/friendly** | with pets | in parks

Did you go for a run yesterday? The weather was perfect.

I ran to the nearest park. Someone let their annoying dog off the leash. I threw him a branch in a huge muddy puddle.

You are always | incredibly | **hostile/friendly** | with pets | in parks

20.

I have no idea what to do about this data breach. My brain is shutting down.

I will deal with it. I'll shut down the server immediately, contact all our customers, and inform the authorities. Please go get some rest.

You are | amazingly | **useless/helpful** | in critical | moments

What should we do about that broken pipe? It'll be a pond in there soon.

I have a workout in fifteen minutes, I never skip one. Just maybe call someone to come fix it, how should I know?

You are | amazingly | **useless/helpful** | in critical | moments

21.

How was work this week? Your girlfriend said she forgot what you look like.

I worked 80 hours this week. We are releasing a new game and I was checking every part for possible glitches. I need it to be perfect.

You surely | deserve a | **pay cut/rise** | for your | efforts

Christine said you didn't come in to work yesterday. Were you sick?

No, I bowled until two in the morning with my college pals and then swam in the canal. I was so drunk I slept for twelve hours.

You surely | deserve a | **pay cut/rise** | for your | efforts

22.

Are you still volunteering? What is it like during the COVID pandemic, are you not concerned?

No, I feel empowered by helping others. I now shop for groceries and medicine for seniors, sew masks for healthcare workers, and make care packages for my neighbors.

You are | the most | **heartless/kind-hearted** | person | I know

Why have you deleted your brother's save in the game? He'll be devastated!

Ah, don't worry. It'll teach him a lesson for taking my X-box without asking. He'll have to start over from the beginning.

You are | the most | **heartless/kind-hearted** | person | I know

23.

I got two speeding tickets just this week! Do you know whether I can pay them online?

Honestly, I've never gotten one. I do my absolute best to never violate any traffic rules.

You always were | such a | **careless/careful** | driver | in the city

How was your trip to IKEA? Did you buy that patio chair?

I got a speeding ticket midway when cops appeared out of nowhere and hit my side-mirror when trying to park. I think I also got a parking ticket.

You always were | such a | **careless/careful** | driver | in the city

24.

I am so lost in life. Do you know what you want to do after graduation?

Well, I decided what I want to be when I was eight and every step just brings me closer to it. I already have a job offer from my dream company.

How does it feel | to be so | **aimless/ambitious** | at your | young age?

Do you have any plans after graduation? I already got a job offer from Google!

I'm a drifter, you know? I guess I will just go with the flow and see where it takes me. I really couldn't care less.

How does it feel | to be so | **aimless/ambitious** | at your | young age?

25.

How did your investment turn out?

Right, you haven't heard the news yet. The start-up I invested in is a massive success and now makes about 100 thousand a month in revenue. I'm rich.

You made | a truly | **stupid/brilliant** | decision | that day

I heard you got a big bonus from our boss. Great job!

Yeah, I thought to save it at first but then I felt lucky and bet it all in the casino. Fortune did not smile on me and I lost everything.

You made | a truly | **stupid/brilliant** | decision | that day

26.

Are you joining us for beers tonight?

A few months ago I kept tabs on how many beers I drink and didn't like the result. So I haven't touched a drop since and do daily workouts instead.

You always were | incredibly | **childish/mature** | in your | behavior

Did you go to that meeting with the rector about funding cuts?

Yes, but I got into a World of Warcraft raid and was 30 minutes late. The rector got angry and didn't let me in.

You always were | incredibly | **childish/mature** | in your | behavior

27.

What are your plans for July?

I joined the local hang gliding club. We have already launched from all the nearest mountains, so we are now planning to glide from an active volcano. I'm so excited!

I had no idea | you were so | **fearful/fearless** | deep | down

We're going camping next month. Are you interested?

Camping is dangerous! What about bears? Bats transmitting rabies? Mosquitoes or other insects that carry Lyme disease and malaria? No, thanks, I'll pass.

I had no idea | you were so | **fearful/fearless** | deep | down

28.

We are planning to go to the mall to have some fun. Would you like to join us?

No, sorry, I am saving money to pursue a Master's in the States. I froze my credit, track all my expenses, and shop only with a list when I need something.

You are | indeed | **wasteful/prudent** | with your | finances

Did you do something fun this weekend before the semester starts?

I did a lot of compulsive shopping. Now I have a Superman costume, a house for my cat, a lightsaber, the complete Discworld series… Oh, and deer antlers.

You are | indeed | **wasteful/prudent** | with your | finances

29.

Why didn't you come to our freshman party this weekend? There were a ton of drinks and we had so much fun!

That's not for me. I run 10 miles every day and have only drank fresh juice and tea for many years.

You have | a super | **harmful/healthy** | lifestyle | for your age

How many cigarettes do you smoke a day? I can't get the smell out of my jacket even after washing it.

About a pack, sometimes more if we have a party with friends. I also really love weed.

You have | a super | **harmful/healthy** | lifestyle | for your age

Fillers

1.

My weekend was so boring. My parents made me declutter the garage. What about you?

I went to a concert of a local band. They played hits from the 80s. The guitar was out of tune and the drummer got drunk right on stage. It was ridiculous.

You really | should have | asked | for your money | back

2.

Is there something on our football team in the morning paper?

Yes, it says they scored 20 points all year. No wonder they are dead last in the league. I mean, how can someone be that bad?

Yeah, you'd think | they could do | better | with a new | couch

3.

I heard you moved houses a month ago. How do you find your new place?

It's a real deal for the money, and close to my kid's school. Sadly, I broke my lawn mower the very first day, so the garden really needs some work.

You can get | a new one | at Home Depot | at a bargain | price

4.

There are a lot of people in the streets. Do you know what's going on?

I believe there is a climate strike today. Last time I checked hundreds of people joined the event on Facebook. It's quite atypical for our small town, isn't it?

I'm glad | environmental | issues | attract so much | attention

5.

Have you painted that picture? It is fabulous! Where did you learn to paint?

I am self-taught really. I used to watch a lot of Bob Ross shows as a kid and decided to make it my hobby. I mostly do watercolors.

You should | try and sell | your artwork | on the internet | or the market

6.

Have you raised enough money to develop that game?

Well, my goal was thirty thousand, and so far I only raised fifteen. I am worried that the project is too niche. I feel like a failure. Maybe I shouldn't have left my job.

You may | want to | ask bloggers | to spread the word | about it

7.

How was your first winter in Manitoba?

I've never been so cold in my entire life. The air was so cold I could hardly breathe. I can't imagine how I am going to stay here for five more years.

I suggest | you try | layering | and thermal | underwear

8.

Did you get that job offer from Visa? I heard you applied.

Yes, they offered me the position of Senior Data Scientist. I am moving to Toronto on the 1st of September. It's a bit sad to leave Vancouver, though.

Congratulations, | I am certain | you will absolutely | love it | there

9.

I can't make myself study for the exams. I have an attention span of a goldfish.

I feel you. Have you tried the Pomodoro technique? It's a huge focus booster for me, and helps manage distractions. Also study groups can help.

Thank you, | that method | sounds | very promising | to me

10.

Where did you learn to code? I am tired of my job and want to try something new.

I took offline courses but those can be really pricey. From what I heard, Codecademy has affordable online options. I could also ask some of my colleagues on Monday.

Money is | tight | at the moment | so I appreciate| the suggestion

11.

You haven't replied to my messages for several days. I was worried something happened.

Sorry, I have no internet connection. From what I understand there are some issues with our phone line. The question is, what am I supposed to do now with this covid thing going on?

Have you | tried | calling | the support | team?

12.

I am definitely a cat person. How can anyone not love these fluffy creatures?

You know, I prefer dogs for many reasons. They love you unconditionally, while cats always treat you like their servant and demand food.

You will | change | your mind | when you meet | my cat

13.

Are you flying somewhere over the Christmas break?

Yeah, I really wanted to visit Germany but my visa got rejected. I forgot to bring my salary slips and account statements. I was so outraged I even sent them an angry email.

There is still | plenty of | time | to try it | again

14.

I am struggling to find something to watch over the weekend. Any suggestions?

I love everything by Jim Jarmusch. He has a very unique style, although some people find the pacing too slow. A good way to start is *Night on Earth*.

That would be | a nice | break | from the Marvel | universe

15.

I saw an amazing blue bird with a crest yesterday. You're a birder! Do you know what that could be?

Must have been a jay, either a Steller's jay or a Blue jay. Steller's jays are much darker and have a black head.

Could you please | send me | the name | so I won't | forget?

A.2 Final stimuli for Chapter 4

n	sentence	condition	familiarity
1	The officer belched during the play.	literal	N/A
2	The school belched out the mediocre student.	metaphor	low
3	The car coasted without any gas.	literal	N/A
4	The athlete coasted through college.	metaphor	high
5	The spider crawled along the edge.	literal	N/A
6	The banker crawled through the contract.	metaphor	low
7	The bees buzzed in the garden.	literal	N/A
8	The headline buzzed in his ears.	metaphor	high
9	The crow flew into the nest.	literal	N/A
10	The sparks flew between the lovers.	metaphor	high
11	The rabbits hopped in the yard.	literal	N/A
12	The insults hopped on her tongue.	metaphor	low
13	The brilliant performer jumped onto the platform.	literal	N/A
14	The detective jumped at the clue.	metaphor	high
15	The happy fan leapt during the goal.	literal	N/A
16	The celebrity leapt at the book deal.	metaphor	high
17	The singer coughed up more blood.	literal	N/A
18	The atm coughed up my card.	metaphor	high
19	The baby birds cried for food.	literal	N/A
20	The building cried for fresh paint.	metaphor	high
21	The lecturer droned for three hours.	literal	N/A
22	The contract droned for many pages.	metaphor	high
23	The lion pounced on its prey.	literal	N/A
24	The newspaper pounced on the story.	metaphor	high
25	Their uncle groaned in the other room.	literal	N/A
26	The inn groaned at the new guests.	metaphor	low
27	The runner raced past the empty lot.	literal	N/A
28	The reader raced through the romance novel.	metaphor	high
29	The kitten purred on the sofa.	literal	N/A
30	The flowers purred in the sunlight.	metaphor	low
31	The sleigh slid down the hill.	literal	N/A
32	The conversation slid into a wall.	metaphor	low
33	The hard candy rattled in the box.	literal	N/A
34	The image rattled in her head.	metaphor	high
35	The snake slithered across the threshold.	literal	N/A
36	The legend slithered across the internet.	metaphor	low
37	The cook stirred up the stew.	literal	N/A
38	The minister stirred up the audience.	metaphor	high
39	The tears streamed down her cheeks.	literal	N/A
40	The pictures streamed through her head.	metaphor	high
41	The vain shopper strutted with pride.	literal	N/A
42	The actor strutted through middle age.	metaphor	low
43	The tired sailor snored in the hammock.	literal	N/A
44	The textbooks snored on the desk.	metaphor	low
45	The old pig snorted at the farmer.	literal	N/A
46	The overhead bin snorted at the large suitcase.	metaphor	low

47	The whale swam through the channel.	literal	N/A
48	The pupil swam through the text.	metaphor	low
49	His hen squawked from her roost.	literal	N/A
50	His knuckles squawked from the punch.	metaphor	low
51	Her bashful suitor stuttered during dinner.	literal	N/A
52	His feet stuttered on the dance floor.	metaphor	high
53	The storm thundered in the distance.	literal	N/A
54	The headache thundered between his temples.	metaphor	high
55	His laundry whirred in the dryer.	literal	N/A
56	His brain whirred in his skull.	metaphor	high
57	The girl tangoed with the instructor.	literal	N/A
58	The garlic tangoed with the ginger.	metaphor	low
59	The hostess whispered to the gentleman.	literal	N/A
60	The cigarettes whispered to the children.	metaphor	low
61	The thin man yelped at the injection.	literal	N/A
62	The suede jacket yelped in the rain.	metaphor	low
63	The mouse wormed through the opening.	literal	N/A
64	The news wormed through the family.	metaphor	low
65	The careless butcher cut herself with a knife.	filler	N/A
66	The hungry chef cooked dinner for herself after work.	filler	N/A
67	The dentist tried to make herself popular with her patients.	filler	N/A
68	The new diplomat drove herself around the capitol.	filler	N/A
69	The famous drummer pictured herself as a singer.	filler	N/A
70	The explorer found herself in uncharted territory.	filler	N/A
71	The cautious jailer armed herself with a gun.	filler	N/A
72	A talented movie director can make herself famous.	filler	N/A
73	The experienced pilot prepared herself for the emergency landing.	filler	N/A
74	The young police officer trained herself to stay calm.	filler	N/A
75	The beauty consultant never let himself get carried away.	filler	N/A
76	The childcare worker fixed himself a cup of tea.	filler	N/A
77	The florist cut himself with the rose thorn.	filler	N/A
78	The young librarian enjoyed himself at the party.	filler	N/A
79	The talkative manicurist made himself laugh.	filler	N/A
80	The adventurous nurse put himself on the list of volunteers.	filler	N/A
81	The stripper made time for himself after the show.	filler	N/A
82	The secretary bought himself a plane ticket.	filler	N/A
83	The hairdresser lit himself a cigarette.	filler	N/A
84	The elementary school teacher hid himself in the closet.	filler	N/A
85	The wealthy queen built herself a castle.	filler	N/A
86	The hungry waitress ordered herself a burger.	filler	N/A
87	The girl scout built herself a fire.	filler	N/A
88	The bachelor cooked himself dinner.	filler	N/A
89	The elderly gentleman fixed himself up for the dance.	filler	N/A
90	The kindly uncle enjoyed himself at Christmas.	filler	N/A
91	Our aerobics instructor gave herself a break.	filler	N/A
92	Our receptionist found herself the center of attention.	filler	N/A
93	My interior decorator prides herself on her work.	filler	N/A
94	A good magician knows how to free himself from a safe.	filler	N/A
95	A successful inventor allows himself to make mistakes.	filler	N/A
96	The dean always gave himself time for a coffee break.	filler	N/A

Appendix B Surveys

B.1 Wilson-Patterson Political Inventory

(modified from Wilson & Patterson, 1968)

Instruction: "Please indicate how much you agree or disagree with the following concepts using a scale provided below"

Response scale: 5-point (1 = strongly disagree to 5 = strongly agree)

Scoring

R = reverse scored

1. school prayer
2. stop all immigration
3. death penalty
4. universal healthcare R
5. gay marriage R
6. right to legal abortion R
7. biblical truth
8. increase welfare spending R
9. increase military spending
10. foreign aid for nations in crisis R
11. lower taxes
12. allow torture of terrorism suspects
13. sex before marriage R
14. gender equity R
15. climate change action R
16. obedience
17. compromise R
18. patriotism
19. gun control
20. free market

Wilson, G. D., & Patterson, J. R. (1968). A new measure of conservatism. British Journal of Social and Clinical Psychology, 7(4), 264–269. https://doi.org/10.1111/j.2044-8260.1968.tb00568.x

B.2 Interpersonal Reactivity Index

(Davis, 1980)

Instruction: “Please indicate how well the following statements describe you using a scale provided below”

Response scale: 5-point (1 = does not describe me very well to 5 = describes me very well)

Scoring

(-) denotes item to be scored in reverse fashion
PT = perspective-taking scale
FS = fantasy scale
EC = empathic concern scale
PD = personal distress scale

1. I daydream and fantasize, with some regularity, about things that might happen to me. (FS)
2. I often have tender, concerned feelings for people less fortunate than me. (EC)
3. I sometimes find it difficult to see things from the "other guy's" point of view. (PT) (-)
4. Sometimes I don't feel very sorry for other people when they are having problems. (EC) (-)
5. I really get involved with the feelings of the characters in a novel. (FS)
6. In emergency situations, I feel apprehensive and ill-at-ease. (PD)
7. I am usually objective when I watch a movie or play, and I don't often get completely caught up in it. (FS) (-)
8. I try to look at everybody's side of a disagreement before I make a decision. (PT)
9. When I see someone being taken advantage of, I feel kind of protective toward them. (EC)
10. I sometimes feel helpless when I am in the middle of a very emotional situation. (PD)
11. I sometimes try to understand my friends better by imagining how things look from their perspective. (PT)
12. Becoming extremely involved in a good book or movie is somewhat rare for me. (FS) (-)
13. When I see someone get hurt, I tend to remain calm. (PD) (-)
14. Other people's misfortunes do not usually disturb me a great deal. (EC) (-)
15. If I'm sure I'm right about something, I don't waste much time listening to other people's arguments. (PT) (-)
16. After seeing a play or movie, I have felt as though I were one of the characters. (FS)
17. Being in a tense emotional situation scares me. (PD)
18. When I see someone being treated unfairly, I sometimes don't feel very much pity for them. (EC) (-)
19. I am usually pretty effective in dealing with emergencies. (PD) (-)
20. I am often quite touched by things that I see happen. (EC)
21. I believe that there are two sides to every question and try to look at them both. (PT)

22. I would describe myself as a pretty soft-hearted person. (EC)
23. When I watch a good movie, I can very easily put myself in the place of a leading character. (FS)
24. I tend to lose control during emergencies. (PD)
25. When I'm upset at someone, I usually try to "put myself in his shoes" for a while. (PT)
26. When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me. (FS)
27. When I see someone who badly needs help in an emergency, I go to pieces. (PD)
28. Before criticizing somebody, I try to imagine how I would feel if I were in their place. (PT)

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. JSAS Catalog of Selected Documents in Psychology, 10, 85.

B.3 Need for Closure Scale

(Webster & Kruglanski, 1994)

Instruction: "Please indicate how much you agree or disagree with the following concepts using a scale provided below"

Response scale: 5-point (1 = strongly disagree to 5 = strongly agree)

Scoring

Reverse-score items 2, 5, 7, 12, 13, 16, 19, 20, 23, 25, 28, 29, 36, 40, 41, and 47.
Sum items 18, 22, 39, 43, and 46 to form a lie score.
Remove the subject if the lie score is greater than 15.
Sum all items except for the above listed lie items to calculate the need for closure score.
Use the top and bottom quartiles to determine high and low need for closure subjects.

If factors are required, use the following scoring system:
Order: 1, 6, 11, 20, 24, 28, 34, 35, 37, 47
Predictability: 5, 7, 8, 19, 26, 27, 30, 45
Decisiveness: 12, 13, 14, 16, 17, 23, 40
Ambiguity: 3, 9, 15, 21, 31, 32, 33, 38, 42
Closed Mindedness: 2, 4, 10, 25, 29, 36, 41, 44

1. I think that having clear rules and order at work is essential for success.
2. Even after I've made up my mind about something, I am always eager to consider a different opinion.
3. I don't like situations that are uncertain.
4. I dislike questions which could be answered in many different ways.
5. I like to have friends who are unpredictable.
6. I find that a well-ordered life with regular hours suits my temperament.
7. I enjoy the uncertainty of going into a new situation without knowing what might happen.
8. When dining out, I like to go to places where I have been before so that I know what to expect.
9. I feel uncomfortable when I don't understand the reason why an event occurred in my life.
10. I feel irritated when one person disagrees with what everyone else in a group believes.
11. I hate to change my plans at the last minute.
12. I would describe myself as indecisive.
13. When I go shopping, I have difficulty deciding exactly what it is I want.
14. When faced with a problem I usually see the one best solution very quickly.
15. When I am confused about an important issue, I feel very upset.
This is an attentiveness check; please indicate "strongly agree".
16. I tend to put off making important decisions until the last possible moment.
17. I usually make important decisions quickly and confidently.

18. I have never been late for an appointment or work.
19. I think it is fun to change my plans at the last moment.
20. My personal space is usually messy and disorganized.
21. In most social conflicts, I can easily see which side is right and which is wrong.
22. I have never known someone I did not like.
23. I tend to struggle with most decisions.
24. I believe orderliness and organization are among the most important characteristics of a good student.
25. When considering most conflict situations, I can usually see how both sides could be right.
26. I don't like to be with people who are capable of unexpected actions.
27. I prefer to socialize with familiar friends because I know what to expect from them.
28. I think that I would learn best in a class that lacks clearly stated objectives and requirements.
29. When thinking about a problem, I consider as many different opinions on the issue as possible.
30. I don't like to go into a situation without knowing what I can expect from it.
This is an attentiveness check; please indicate "strongly disagree".
31. I like to know what people are thinking all the time.
32. I dislike it when a person's statement could mean many different things.
33. It's annoying to listen to someone who cannot seem to make up his or her mind.
34. I find that establishing a consistent routine enables me to enjoy life more.
35. I enjoy having a clear and structured mode of life.
36. I prefer interacting with people whose opinions are very different from my own.
37. I like to have a plan for everything and a place for everything.
38. I feel uncomfortable when someone's meaning or intention is unclear to me.
39. I believe that one should never engage in leisure activities.
40. When trying to solve a problem I often see so many possible options that it's confusing.
41. I always see many possible solutions to problems I face.
42. I'd rather know bad news than stay in a state of uncertainty.
43. I feel that there is no such thing as an honest mistake.
44. I do not usually consult many different options before forming my own view.
45. I dislike unpredictable situations.
46. I have never hurt another person's feelings.
47. I dislike the routine aspects of my work (studies).

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. Journal of Personality and Social Psychology, 67(6), 1049–1062. https://doi.org/10.1037/0022-3514.67.6.1049

B.4 Language Background Questionnaire for Chapter 4

1. How old are you?
2. Which gender do you identify with the most?
   - Male
   - Female
   - Non-binary
   - Trans
   - Rather not say
   - Other
3. Do you identify as American?
   - Yes
   - No
4. Are you bilingual (native in two languages)?
   - Yes
   - No
5. How old were you when you started learning English?
6. How many foreign languages do you speak (at any level of proficiency, please don't count your native language or languages)?
   - 0
   - 1
   - 2
   - 3
   - 4 or more
7. Do any of your parents have a first language other than English?
   - No
   - Yes, one of them
   - Yes, both of them
8. Have you ever had a close friend whose first language was not English?
   - No
   - Yes, one
   - Yes, two or more
9. Do you have any relatives whose first language is not English?
   - No
   - Yes, one
   - Yes, two or more
10. How ethnically diverse was the environment in which you grew up?
    - Zero
    - Some diversity
    - Fair diversity
    - High diversity
    - Very high diversity
11. How linguistically diverse (people with foreign accents) was the environment in which you grew up?

- Zero
- Some diversity
- Fair diversity
- High diversity
- Very high diversity

12. How ethnically diverse is your current school/university or workplace?
    - Zero
    - Some diversity
    - Fair diversity
    - High diversity
    - Very high diversity
13. How linguistically diverse (people with foreign accents) is your current school/university or workplace?
    - Zero
    - Some diversity
    - Fair diversity
    - High diversity
    - Very high diversity
14. How often do you interact with native speakers of English?
    - Very infrequently
    - Infrequently
    - Average
    - Frequently
    - Very frequently
15. How often do you interact with non-native speakers of English?
    - Very infrequently
    - Infrequently
    - Average
    - Frequently
    - Very frequently

Based partially on:

Dewaele, J.-M., & McCloskey, J. (2014). Attitudes toward foreign accents among adult multilingual language users. *Journal of Multilingual and Multicultural Development, 36*(3), 221–238. https://doi.org/10.1080/01434632.2014.909445

Dewaele, J.-M., & Wei, L. (2012). Is multilingualism linked to a higher tolerance of ambiguity? Bilingualism: Language and Cognition, 16(01), 231–240. https://doi.org/10.1017/S1366728912000570

B.5 Cognitive Reflection Test

(adapted from Frederick, 2005; Thomson & Oppenheimer, 2016)

Math:

1. If three elves can wrap three toys in an hour, how many elves are needed to wrap six toys in 2 hours?

(correct: 3 elves, heuristic: 6)

2. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

(correct: 47 days; heuristic: 24)

3. In an athletics team, tall members are three times more likely to win a medal than short members. This year the team has won 60 medals so far. How many of these have been won by short athletes?

(correct: 15 medals; heuristic: 20)

Verbal:

4. A farmer had 15 sheep and all but 8 died. How many are left?

(correct: 8; heuristic: 7)

5. If you're running a race and you pass the person in second place, what place are you in?

(correct: 2 place; heuristic: 1)

6. How many cubic feet of dirt are there in a hole that is 3' deep x 3' wide x 3' long?

(correct: 0 feet of dirt; heuristic: 27)

Decoy:

7. A cargo hold of a ship had 500 crates of oranges. At the ship's first stop, 100 crates were unloaded. At the second stop, 200 more were unloaded. How many crates of oranges were left after the second stop?

8. Sara, Emma, and Sophia embark on a river trip. Each of them brings one supply item for the trip: a kayak, a cooler of sandwiches, and a bag of apples. Sara brought the apples and Emma didn't bring anything edible. What did Sophia bring?

9. An expedition on a mountain climbing trip was traveling with eleven horse packs. Each horse can carry only three packs. How many horses does the expedition need?

10. A mechanic shop had five silver cars, two red cars, and one blue car in the garage. During the day, three silver cars and one red car were picked up, and one black car was dropped off. How many silver cars were in the garage at the end of the day?

Frederick, S. (2005). Cognitive Reflection and Decision Making. Journal of Economic Perspectives, 19(4), 25–42. https://doi.org/10.1257/089533005775196732

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. Judgment and Decision Making, 11(1), 99–113. https://doi.org/10.1017/S1930297500007622

B.6 Accent Attitudes Scale

(modified from Contemori & Tortajada, 2020; Weatherholtz et al., 2014)

Instruction: "please indicate how much you agree with each statement"

Response scale: 10-point (1 = do not agree at all and 10 = agree completely)

Scoring

R = reverse scored

1. It bothers me when one doesn't speak English properly.
2. I enjoy hearing accents from different places. R
3. Speaking well is important to me.
4. Accent is an important part of one's self-presentation.
5. People with strong accents are just as likely to be smart as people without accents. R
6. When I hear someone speak with an accent that is different from my own, I expect to have difficulty understanding them.
7. I would work well in a professional or school environment with a person who speaks with a foreign accent. R
8. I think that the United States should be an English-monolingual country.
9. When I hear a speaker of another language speaking English, I want to hear it accurately spoken.

Contemori, C., & Tortajada, F. (2019). The use of social–communicative cues to interpret ambiguous pronouns: Bilingual adults differ from monolinguals. Applied Psycholinguistics, 1–27. https://doi.org/10.1017/s0142716419000407

Weatherholtz, K., Campbell-Kibler, K., & Jaeger, T. F. (2014). Socially-mediated syntactic alignment. Language Variation and Change, 26(3), 387–420. https://doi.org/10.1017/S0954394514000155

B.7 Language Background Questionnaire for Chapter 5

1. How old are you?
2. Which gender do you identify with the most?
   - Male
   - Female
   - Non-binary
   - Trans
   - Rather not say
   - Other
3. Do you identify as Canadian?
   - Yes
   - No
4. Do any of your parents have a first language other than English?
   - No
   - Yes, one of them
   - Yes, both of them
5. Have you ever had a close friend whose first language was not English?
   - No
   - Yes, one
   - Yes, two or more
6. Do you have any relatives whose first language is not English?
   - No
   - Yes, one
   - Yes, two or more
7. How ethnically diverse was the environment in which you grew up?
   - Zero
   - Some diversity
   - Fair diversity
   - High diversity
   - Very high diversity
8. How linguistically diverse (people with foreign accents) was the environment in which you grew up?
   - Zero
   - Some diversity
   - Fair diversity
   - High diversity
   - Very high diversity
9. How ethnically diverse is your current school/university or workplace?
   - Zero
   - Some diversity
   - Fair diversity
   - High diversity
   - Very high diversity

10. How linguistically diverse (people with foreign accents) is your current school/university or workplace?
- Zero
- Some diversity
- Fair diversity
- High diversity
- Very high diversity

11. How often do you interact with native speakers of English?
- Very infrequently
- Infrequently
- Average
- Frequently
- Very frequently

12. How often do you interact with non-native speakers of English?
- Very infrequently
- Infrequently
- Average
- Frequently
- Very frequently

Additionally in Experiment 2:

13. Have you lived in Alberta for most of your life?
- Yes
- No

14. How old were you when you started learning English?

15. Are you bilingual (native in two languages)?
- Yes
- No

16. How many foreign languages do you speak (at any level of proficiency, please don't count your native language or languages)?
- 0
- 1
- 2
- 3
- 4 or more

Based partially on:

Dewaele, J.-M., & McCloskey, J. (2014). Attitudes toward foreign accents among adult multilingual language users. *Journal of Multilingual and Multicultural Development, 36*(3), 221–238. https://doi.org/10.1080/01434632.2014.909445

Dewaele, J.-M., & Wei, L. (2012). Is multilingualism linked to a higher tolerance of ambiguity? Bilingualism: Language and Cognition, 16(01), 231–240. https://doi.org/10.1017/S1366728912000570

Appendix C Supplementary analyses

C.1 The Implicit Association Test procedure for Chapter 4

We conducted an IAT test partially based on Roessel et al. (2018) and Lee (2015). As target concepts, we used "AMERICAN ENGLISH" and "FOREIGN ACCENT". Both of these terms exclusively refer to language and thus cannot be confused with nationality, which helps avoid potential confounds (K. A. Lane et al., 2007). The attribute concepts were "GOOD" and "BAD". Attribute words were selected to represent the dimension of competence: COMPETENT/INCOMPETENT, KNOWLEDGEABLE/IGNORANT, INTELLIGENT/DULL, EDUCATED/UNEDUCATED, SMART/STUPID. Attribute words did not differ significantly between the two categories either in frequency or length according to Student's *t*-tests (*p*s > .5).

Materials and procedure

In line with common practice, the test consisted of seven blocks. The participants were instructed to sort the terms that appeared in the center of the screen with the categories in the upper right and left corners of the screen by pressing the 'A' or 'L' keys. In case of an incorrect response, a red cross appeared below the stimulus until a correct answer was given. The participants were instructed to go as quickly as possible while being accurate. Target words and the categories they should be sorted into were presented before the target-practice block (Block 1) and attribute words with the categories they should be sorted into were presented before the attribute-practice block (Block 2). The order of the Blocks 3-4 and 6-7 was counterbalanced between participants so that half of them got the compatible block first (AMERICAN ENGLISH OR GOOD/FOREIGN ACCENT OR BAD) and the other half got the incompatible block first (FOREIGN ACCENT OR GOOD/AMERICAN ENGLISH OR BAD). Only the starting positions for categories in Block 1 were counterbalanced so that the initial combined block was also counterbalanced between subjects. The attribute block (Block 2) was not counterbalanced.

Data analysis

An updated scoring algorithm by Greenwald et al. (2003) was used. According to the algorithm, there are at least three different ways to calculate error penalty: 1) error trials = block mean of correct trials + 600 ms 2) error trials = block mean of correct trials + 2 SD 3) error trials = built-in error penalty (when participants have to correct their error themselves; latency to correct response is used). Since our participants were shown a red cross following an error and had to correct their response, we used the third method.

Results

Forty-nine participants had the congruent block first and 36 participants had the incongruent block first. This imbalance was due to the fact that the block order was randomly assigned in an online experiment. Four data points were removed due to exceeding 10 000 ms. None of the participants had more than 10% of responses below 300 ms (the highest was 1.7%), so all the participants could be retained. Mean IAT score (where zero is the absence of bias in favor of any accent and scores above zero indicate a bias against foreign accent and in favor of native accent) was 0.51 (SD = 0.4, range -0.5 to 1.6). Visually, the scores were normally distributed. A one-sample *t*-test on IAT scores against zero showed that the bias against foreign accents was statistically significant (*t* =

12.5, $p < 0.001$) and the effect size was large (Cohen's $d = 1.36$). A simple linear model of IAT scores versus block order showed a significant effect ($t = -2.19$, $p = 0.031$). IAT scores were larger in the compatible first versus incompatible first block order (M = 0.58 vs. 0.41).

Hence, in the IAT focused on general category associations, foreign-accented speakers were implicitly considered less knowledgeable than native speakers.

References

Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and Using the Implicit Association Test: IV: What We Know (So Far) about the Method. In *Implicit measures of attitudes.* (pp. 59–102). The Guilford Press.

Lee, R. (2015). Implicit Associations with Welsh in Two Educational Contexts. *York Papers in Linguistics*, *2*(14), 81–105.

Roessel, J., Schoel, C., & Stahlberg, D. (2018). What's in an accent? General spontaneous biases against nonnative accents: An investigation with conceptual and auditory IATs: General biases against nonnative accents. *European Journal of Social Psychology*, *48*(4), 535–550. https://doi.org/10.1002/ejsp.2339

C.2 Individual differences analysis for Chapter 4

Accent attitudes scale. There was a marginally significant main effect of AAS scores (β = -1.17, CI = -2.38 – 0.05, t = -1.88, p = 0.060) qualified by a significant three-way interaction with Speaker and unfamiliar metaphors (β = 0.65, CI = 0.17 – 1.13, t = 2.67, p = 0.008). The plot of the three-way interaction is provided below.

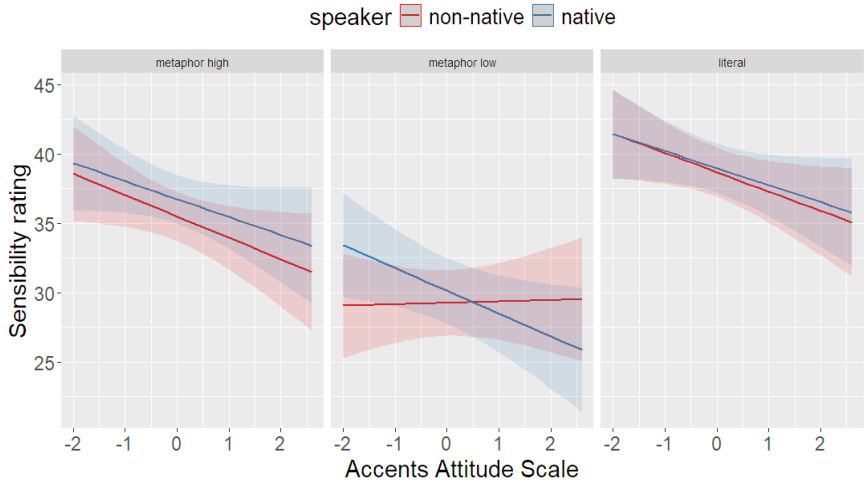

Figure C2.1. Predicted values (marginal effects) of the interaction between speaker condition, metaphor condition, and accent attitudes, with sensibility rating as a dependent variable. Higher scores indicate higher explicit bias against non-standard accents.

IAT. There was a marginally significant main effect of IAT scores (β = -3.14, CI = -2.38 – 0.05, t = -1.88, p = 0.064) qualified by a two-way interaction with unfamiliar metaphors (β = -1.93, CI = -3.20 – -0.66, t = -2.97, p = 0.003). The plot of the two-way interaction is provided below.

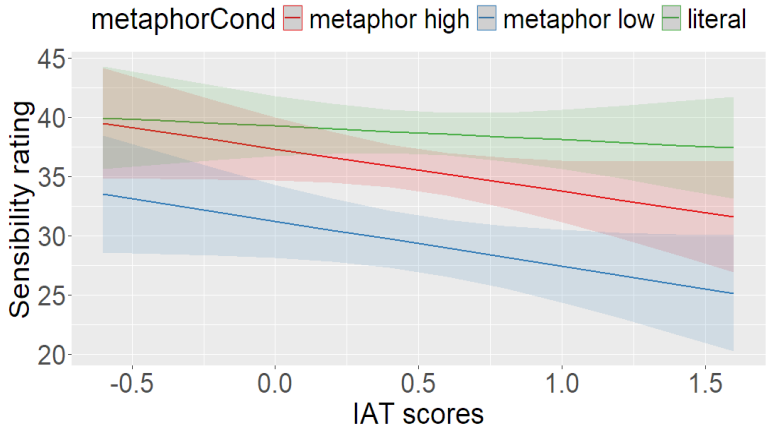

Figure C2.2. Predicted values (marginal effects) of the interaction between metaphor condition, and IAT scores, with sensibility rating as a dependent variable. Higher scores indicate higher implicit bias against foreign accents.

C.3 Generalized additive mixed models for acoustic measures for Chapter 5

Asterisks indicate statistical significance (. p < 0.1, *p < 0.05, **p < 0.01 and ***p < 0.001).

Duration (in ms)

Predictors	*Estimates*	*CI*	*p*	
Intercept	2629.77	2356.05 – 2903.48	<0.001	***
accent [foreign]	779.39	661.32 – 897.46	<0.001	***
truth value [false]	-137.81	-255.88 – -19.74	0.022	*
truth value [true]	-146.81	-264.88 – -28.74	0.015	*
accent [foreign] x truth value [false]	-34.71	-201.69 – 132.27	0.683	
accent [foreign] x truth value [true]	-54.56	-221.54 – 112.42	0.522	
Smooth term (triplet number)			<0.001	***
Smooth term (voice)			<0.001	***
Observations	1080			
R^2	0.346			

Pitch (in Hz)

Predictors	*Estimates*	*CI*	*p*	
Intercept	208.91	179.95 – 237.87	<0.001	***
accent [foreign]	17.42	14.01 – 20.84	<0.001	***
truth value [false]	2.91	-0.51 – 6.32	0.095	
truth value [true]	5.58	2.16 – 8.99	0.001	**
accent [foreign] x truth value [false]	-1.78	-6.61 – 3.05	0.470	
accent [foreign] x truth value [true]	-5.82	-10.64 – -0.99	0.018	*
Smooth term (triplet number)			0.223	
Smooth term (voice)			<0.001	***
Observations	1080			
R^2	0.507			

Speech rate (in nsyll/s)

Predictors	*Estimates*	*CI*	*p*	
Intercept	4.79	4.65 – 4.93	<0.001	***
accent [foreign]	-0.98	-1.10 – -0.86	<0.001	***
truth value [false]	-0.01	-0.13 – 0.11	0.869	
truth value [true]	0.03	-0.09 – 0.15	0.670	
accent [foreign] x truth value [false]	0.07	-0.10 – 0.24	0.429	
accent [foreign] x truth value [true]	-0.00	-0.17 – 0.16	0.959	
Smooth term (triplet number)			0.064	
Smooth term (voice)			0.002	**
Observations	1080			
R^2	0.409			