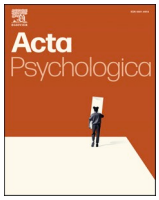
Acta Psychologica 254 (2025) 104853

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy



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

Veranika Puhacheuskaya *, Juhani Järvikivi

University of Alberta, Edmonton, Alberta, Canada

ARTICLE INFO

Keywords:
Linguistic expectations
Non-native accents
Metaphors
Pragmatics
Individual differences

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.

1. Introduction

Imagine receiving a mysterious email with just one sentence: "Conscience is a burrowing mole".[1] 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., 2008a; Van Berkum et al., 2009). 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 et al., 2008a). 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

* Corresponding author.
E-mail addresses: puhacheu@ualberta.ca (V. Puhacheuskaya), jarvikivi@ualberta.ca (J. Järvikivi).

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

https://doi.org/10.1016/j.actpsy.2025.104853
Received 10 October 2024; Received in revised form 27 January 2025; Accepted 20 February 2025
Available online 6 March 2025

in shaping listener expectations. Broadly speaking, accent is a person's unique manner of pronunciation (Giles, 1970) 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 (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 and 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, Brouwer, & Foucart, 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 hypothesize 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 and 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 and 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 and 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 and 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., 2009).

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.

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, Brouwer, & Foucart, 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 hypothesize 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 and 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 and 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.

2. Methods

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.

2.2. Stimuli

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[2] 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 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 ($ts > 4$, $ps < 0.001$).

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)

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. All the stimuli are available on Open Science Framework.

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 and 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[3] 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 and 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

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

[3] 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.

Table 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)

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 pseudorandomized so that no more than three repetitions of the Metaphor/Literal/Filler condition could occur in a row. Fig. 1 shows an example of a trial from a native and a non-native 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 full list of questions for each survey is available in Supplementary materials. 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.

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).

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 and 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.

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

Implicit Association Test (IAT). The detailed description and the results of the test are available in Supplementary materials 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 min to complete.

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 Open Science Framework. 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. The detailed description of the calculation of individual difference scores, their distribution, and reliability is available in Supplementary material, together with the full correlation matrix. 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.

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

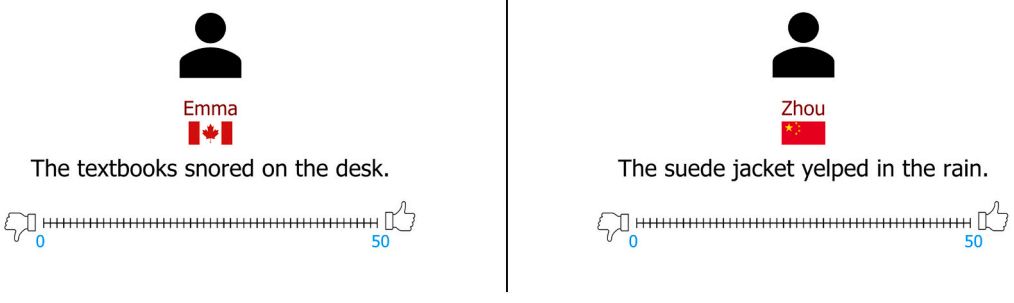


Fig. 1. An example of a trial with an unfamiliar metaphor from a native (on the left) and a non-native block.

Table 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)

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.

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, Brouwer, and Foucart (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 as well.

Ratings and RTs. Mean sensibility ratings and response times for every experimental condition are reported in Table 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.

For linear-mixed models, both factorial predictors were sum-coded (Speaker and Metaphor Condition). The results from the main model with sensibility ratings as a dependent variable and only the manipulated and control variables as predictors are reported in Table 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. Naturalness affected ratings in a predicted direction: more natural sentences were rated as more sensible than less natural ones.

Table 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 $R^2 = 0.548$. Asterisks indicate statistical significance (*p < 0.05, **p < 0.01 and ***p < 0.001).

Predictors	*β*	*CI*	*p*
(Intercept)	13.30	9.97–16.63	<0.001***
Non-native speaker	−0.40	−0.72 to −0.08	0.013*
Familiar metaphors	1.24	0.46–2.02	0.002**
Unfamiliar metaphors	−5.18	−6.68 to −3.68	<0.001***
Naturalness	4.58	3.92–5.23	<0.001***
Imageability	−0.06	−0.66–0.53	0.834
Non-native speaker × familiar metaphors	−0.22	−0.70–0.25	0.358
Non-native speaker × unfamiliar metaphors	−0.03	−0.51–0.44	0.898

Table 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 $R^2 = 0.433$. Asterisks indicate statistical significance (*p < 0.05, **p < 0.01 and ***p < 0.001).

Predictors	*β*	*CI*	*p*
(Intercept)	72.15	69.10–75.20	<0.001***
Non-native speaker	0.40	0.12–0.68	0.006**
Familiar metaphors	1.09	0.43–1.76	0.001**
Unfamiliar metaphors	−2.24	−3.52 to −0.95	<0.001***
Naturalness	−2.78	−3.34 to −2.22	<0.001***
Imageability	0.08	−0.43–0.59	0.768
Non-native speaker × familiar metaphors	0.26	−0.15–0.68	0.212
Non-native speaker × unfamiliar metaphors	−0.16	−0.58–0.26	0.442

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. For familiar metaphors, 23 % of responses were at ceiling, whereas 15 % of unfamiliar metaphors were at floor. The ceiling effects may have thus reduced the speaker effects for the literal condition by compressing the variability of responses.

The model with reaction times as a predictor is provided in Table 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.

We will now go over the individual differences.

Language 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

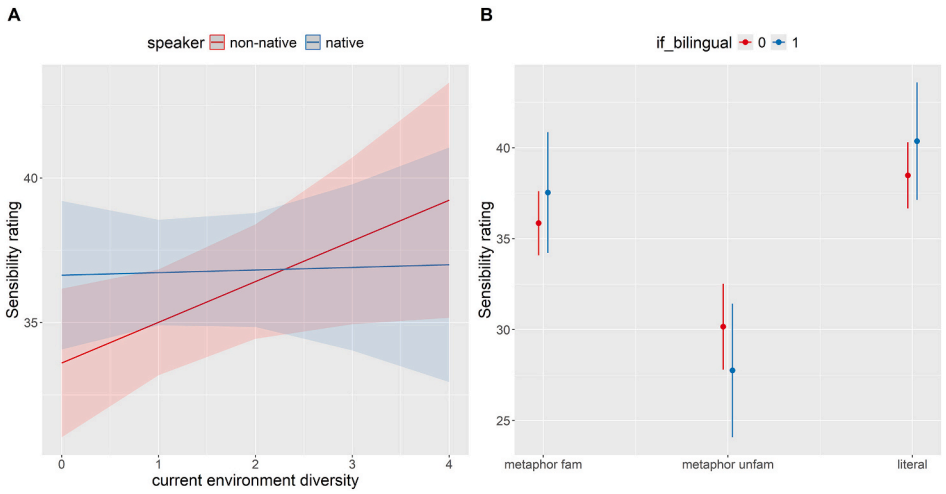


Fig. 2. Predicted values (marginal effects) of the interaction between current environment diversity score and speaker (A) and between being bilingual and metaphor condition (B). The dependent variable is sensibility rating.

significant improvement in the model's fit ($\chi^2 = 10.85$, $p = 0.093$). Next, we examined the participant's self-reported childhood language diversity (described as the number of people around with foreign accents), but it also failed to improve the model ($\chi^2 = 6.75$, $p = 0.344$). Including the self-reported language diversity of the participant's current environment marginally improved the model ($\chi^2 = 11.93$, $p = 0.063$). 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. 2A, 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.[4]

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 ($\chi^2 = 24.0$, $p < 0.001$). 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: $\beta = -2.79$, $CI = -4.03$ to -1.55, $t = -4.41$, $p < 0.001$). As shown in Fig. 2B, 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 % of our participants self-identified as bilingual, so this effect is worth replicating with a larger participant sample.

Cognitive Reflection Test. CRT scores significantly improved the base model ($\chi^2 = 82.72$, $p < 0.001$). 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 ($\beta = -1.14$, $CI = -1.44$ to -0.84, $t = -7.45$, $p < 0.001$). Both interactions are shown in Fig. 3. As can be seen in Fig. 3A, ratings for the non-native speaker consistently improved with higher CRT scores, whereas ratings for the native speaker remained unaffected. Fig. 3B 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's Conservatism.[5] W-P scores significantly improved the baseline model ($\chi^2 = 22.38$, $p = 0.001$). There was a main effect of W-P scores ($\beta = -1.44$, $CI = -2.65$ to -0.24, $t = -2.35$, $p = 0.021$) 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. 4A 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.

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 Supplementary material only.

Big-Five Inventory. Openness did not improve the model's fit ($\chi^2 = 5.78$, $p = 0.449$), but Extraversion did ($\chi^2 = 33.3$, $p < 0.001$). 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. 4B, while native unfamiliar metaphors were rated similarly regardless of Extraversion, non-native novel metaphors were rated significantly higher the more extraverted the participant was.

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.,

[4] 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 environment diversity.

[5] Accents Attitudes Scale scores also significantly improved the base model ($\chi^2 = 14.44$, $p = 0.025$), 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 (42,234 vs. 42,242) and BIC (42,346 vs. 42,354). 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 Supplementary material).

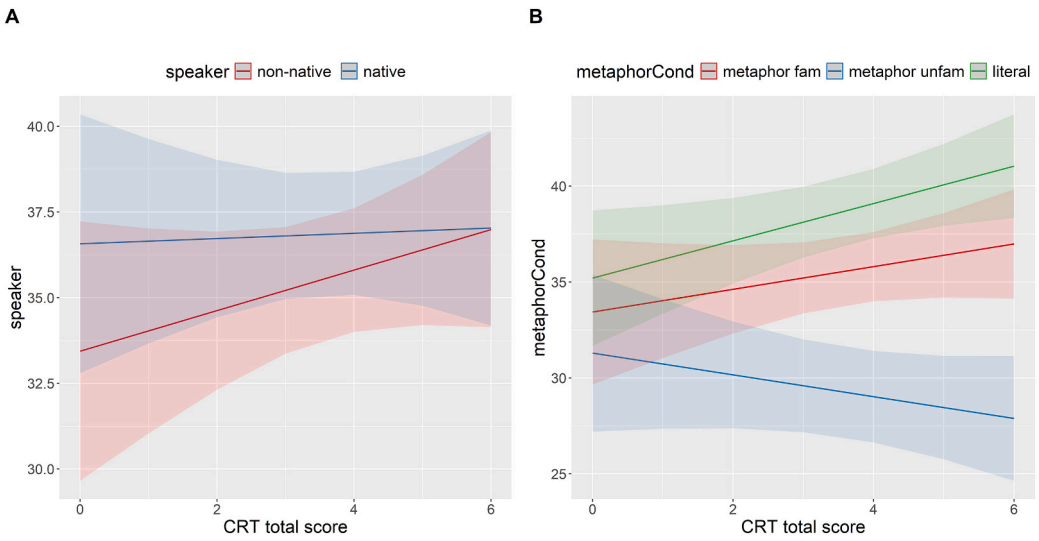


Fig. 3. Predicted values (marginal effects) of the interaction between CRT scores and speaker condition (A) and between CRT scores and metaphor condition (B), with sensibility rating as a dependent variable. Higher scores indicate higher cognitive reflection.

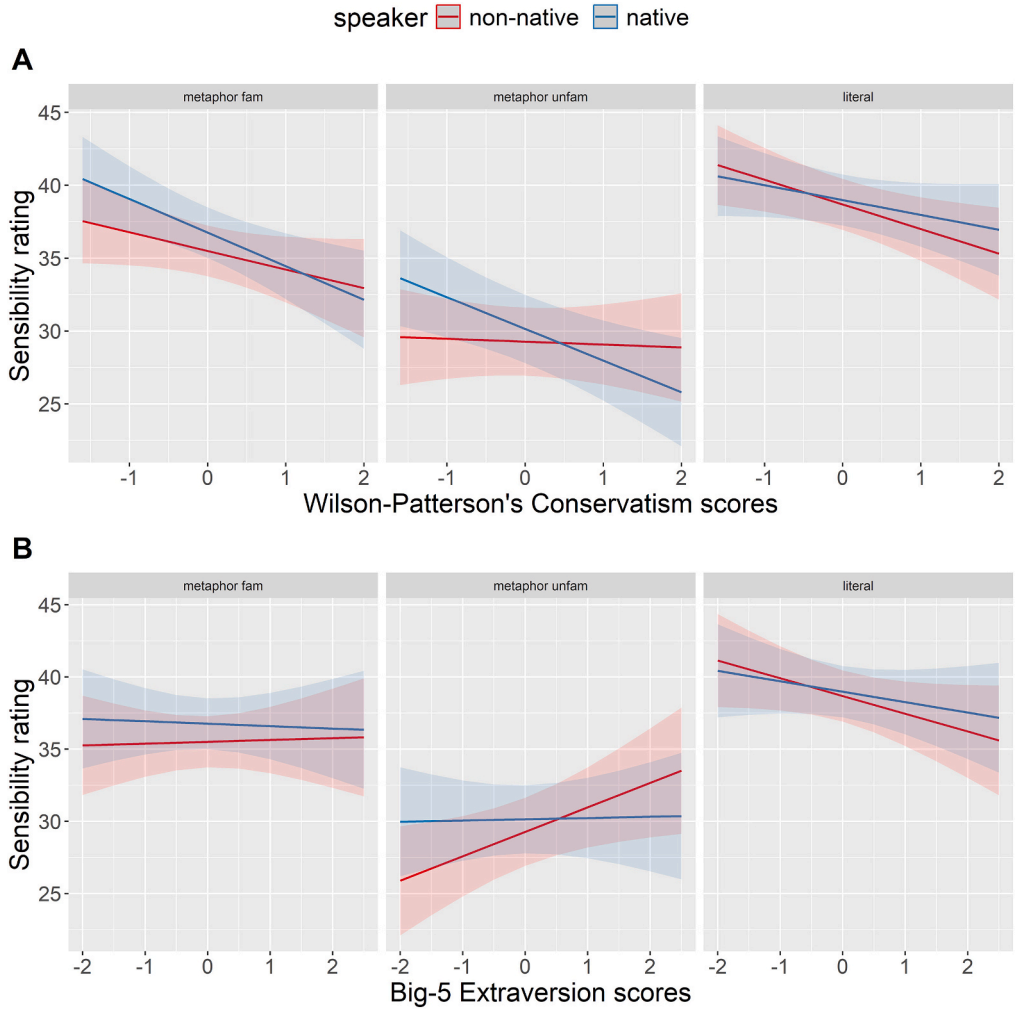


Fig. 4. Predicted values (marginal effects) of the interaction between speaker condition, metaphor condition, and political orientation (A) and Big-5 Trait Extraversion scores (B), with sensibility rating as a dependent variable. Higher scores indicate higher conservatism and higher extraversion.

2018; Pélissier & Ferragne, 2022; Van Berkum et al., 2008a) and listener (Eekhof et al., 2021; Hammond-Thrasher & Järvikivi, 2023; Hubert Lyall, 2019; Mayn & Demberg, 2022; Puhacheuskaya & Järvikivi, 2022; Van Berkum et al., 2009) 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*. In fact, 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 cognitively taxing. These findings contribute 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 et al., 2008b). 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, and 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 in the model 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 eliminate any effects of cognitive load due to speech disfluency and thus better isolate the effects triggered by foreignness and expected 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.[6] 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, 2016a, 2016b). 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

[6] We thank the anonymous reviewer for bringing up this issue.

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 our 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).[7] 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 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 % conservative and 50 % liberal participants. We obtained a very good distribution of scores in the Wilson-Patterson's 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 this 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

[7] It is noteworthy that CRT scores were highly correlated with IQ measures in Mayn and 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.

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.

CRediT authorship contribution statement

Veranika Puhacheuskaya: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Juhani Järvikivi:** Writing – review & editing, Supervision, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are very grateful to the journal editor and two anonymous reviewers for their helpful comments on earlier drafts of this manuscript. This research was supported by a Social Sciences and Humanities Research Council of Canada Partnership Grant (Words in the World, 895-2016-1008).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.actpsy.2025.104853.

Data availability

All experiment stimuli, raw data, and R scripts used for data preprocessing and analysis are available on Open Science Framework at https://osf.io/yf75c/.

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*, Article 105007. https://doi.org/10.1016/j.bandl.2021.105007

Arthur, W., & Day, D. V. (1994). Development of a short form for the Raven advanced progressive matrices test. *Educational and Psychological Measurement, 54*(2), 394–403. https://doi.org/10.1177/0013164494054002013

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

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

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

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

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

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.

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), Article e69173. https://doi.org/10.1371/journal.pone.0069173

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

Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Lang & Ling Compass, 8*(2), 37–50. https://doi.org/10.1111/lnc3.12061

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

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Caffarra, S., Michell, E., & Martin, C. D. (2018). The impact of foreign accent on irony interpretation. *PLoS ONE, 13*(8), Article 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

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

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

Dewaele, J.-M., & McCloskey, J. (2015). Attitudes towards 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

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

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

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*, Article 698986. https://doi.org/10.3389/fpsyg.2021.698986

Fairchild, S., Mathis, A., & Papafragou, A. (2020). Pragmatics and social meaning: Understanding under-informativeness in native and non-native speakers. *Cognition, 200*, Article 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

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*, Article 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.

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

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

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., 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

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.

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

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

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*, Article 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).

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

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

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

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

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

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

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*, Article 699071. https://doi.org/10.3389/fpsyg.2021.699071

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

John, O. P., & Srivastava, S. (1991). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin, & O. P. John (Eds.), *Vol. 2. Handbook of personality: Theory and research* (pp. 102–138). Guilford Press.

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., 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

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).

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

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

Lane, H. (1963). Foreign accent and speech distortion. *The Journal of the Acoustical Society of America, 35*(4).

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.

Lorenzoni, A., Pagliarini, E., Vespignani, F., & Navarrete, E. (2022). Pragmatic and knowledge range lenience towards foreigners. *Acta Psychologica, 226*, Article 103572. https://doi.org/10.1016/j.actpsy.2022.103572

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

Mayn, A., & Demberg, V. (2022). Individual differences in a pragmatic reference GAME. In *Proceedings of the 44th annual conference of the cognitive science society*.

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

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

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

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

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

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … 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

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

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

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

R Core Team. (2020). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/.

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

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

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

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

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

Souza, A. L., & Markman, A. B. (2013). Foreign accent does not influence cognitive judgments. In *Proceedings of the annual meeting of the cognitive science society* (p. 35).

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

Stieger, S., & Reips, U.-D. (2016). A limitation of the cognitive reflection test: Familiarity. *PeerJ, 4*, Article e2395. https://doi.org/10.7717/peerj.2395

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

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

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

van Berkum, J. J. A. (2018). In S.-A. Rueschemeyer, & M. G. Gaskell (Eds.), *Language comprehension, emotion, and sociality* (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. (2009). 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. (2008a). 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. (2008b). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*(4), 580–591. https://doi.org/10.1162/jocn.2008.20054

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

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

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. In *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

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

Wilson, G. D., & Patterson, J. R. (1968). A new measure of conservatism. *British Journal of Social & Clinical Psychology, 7*(4), 264–269. https://doi.org/10.1111/j.2044-8260.1968.tb00568.x

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