Cognition 181 (2018) 80–92

Contents lists available at ScienceDirect

Cognition

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

Original Articles

Sins of omission are more likely to be forgiven in non-native speakers☆

Sarah Fairchild*, Anna Papafragou

Department of Psychological & Brain Sciences, University of Delaware, USA[1]

ARTICLE INFO

Keywords:
Non-native speech
Bilingualism
Pragmatics
Informativeness
Scalar implicature
Expectations

ABSTRACT

Utterances produced by foreign-accented speakers are often judged as less credible, more vague, and more difficult to understand compared to those produced by native speakers. Some theoretical accounts argue that listeners have different expectations about the speech of non-native speakers. Other accounts argue that non-native speech is processed differently to the extent that a foreign accent taxes intelligibility and introduces additional processing load. Here we test the role of expectations for the processing of native vs. non-native speech in written texts where accents cannot be directly perceived (and thus affect processing load). In Experiment 1, native comprehenders gave higher ratings to the meaning of under-informative sentences ("Some people have noses with two nostrils") when they believed that the sentences were produced by non-native compared to native speakers. This difference was larger the more likely individual participants were to interpret under-informative sentences pragmatically (as opposed to logically). In Experiment 2, the tendency to forgive sins of information omission was shown to depend on the presumed L2 proficiency of non-native speakers. Experiment 3 replicated and extended the major finding. Since intelligibility of the sentences was identical across types of speakers, these findings provide support for the role of expectations for non-native speech comprehension, as well as for broader models of language processing that argue for a role of speaker identity.

1. Introduction

There are over 51 million Americans who speak a language other than English at home (U.S. Census Bureau, 2011). Worldwide, it is estimated that about half of the population is bilingual (Grosjean, 2010). Psycholinguistic research on bilingualism has flourished in recent decades, with investigations focusing on the mechanisms allowing bilinguals to switch effortlessly between their two languages (e.g., Poplack, 1980; Clyne, 1987; Milroy & Muysken, 1995; Myers-Scotton & Jake, 2000; Moreno, Federmeier, & Kutas, 2002), the organization of the bilingual mental lexicon (e.g., De Bot & Schreuder, 1993; Kroll & Stewart, 1994; Green, 1998; Wei, 2002; Pavlenko, 2009), and the potential cognitive advantages of bilingualism (e.g., Bialystok, Craik, Klein, & Viswanathan, 2004; Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009; Prior & MacWhinney, 2010; Hilchey & Klein, 2011; Sebastián-Gallés, Albareda-Castellot, Weikum, & Werker, 2012; Paap & Greenberg, 2013). However, the prevalence of bilingualism has consequences that extend beyond the bilingual individuals themselves. Monolinguals frequently interact with individuals who speak more than one language, many of whom are non-native speakers of the language. Because second language phonology is notoriously difficult for adults to acquire (Flege, Yeni-Komshian, & Liu, 1999; Golestani & Zatorre, 2009; Piske, MacKay, & Flege, 2001), many of these non-native speakers speak with a foreign accent. A recent line of research exploring how speech from non-native speakers is processed by native listeners suggests that a foreign accent impacts communication in a number of ways.

Most obviously, foreign-accented speech poses a challenge for intelligibility. Non-native speakers may not be able to properly produce the phonemic inventory of their second language, causing listeners to have difficulty in comprehension. Indeed, participants are slower to process sentences uttered by non-native speakers, and rate such foreign-accented sentences as less comprehensible (Munro & Derwing, 1995). Unsurprisingly, research also demonstrates that word identification is initially impaired by non-native speech in both adult (Bent & Bradlow, 2003; Clarke & Garrett, 2004) and infant (van Heugten & Johnson, 2014) native listeners, although listeners are generally able to quickly adapt to a foreign accent (Baese-Berk, Bradlow, & Wright, 2013; Clarke & Garrett, 2004; van Heugten & Johnson, 2014).

Foreign-accented speech has broader effects on language comprehension: a recent study using Event-Related Potentials (ERPs) found

☆ We thank Hannah Schwartz and Harry Rowe for assistance with data collection and stimuli creation. This work was partially supported by a General University Research Fund (UD) and NSF BCS grant #1632849.

* Corresponding author at: Department of Psychological & Brain Sciences, University of Delaware, 401 Wolf Hall, Newark, DE 19716, USA.
E-mail address: sarahf@udel.edu (S. Fairchild).
[1] S.F. is now with Core Human Factors, Inc.

https://doi.org/10.1016/j.cognition.2018.08.010
Received 1 December 2016; Received in revised form 12 August 2018; Accepted 17 August 2018
Available online 24 August 2018

that neural signatures differed in response to native and non-native (Turkish-accented) speech errors in Dutch (Hanulíková, Van Alphen, Van Goch, & Weber, 2012). Participants showed the typical P600 response to syntactic violations (grammatical gender errors) and N400 to semantic violations (e.g., the Dutch translation of "It was very cold last night, so I put a thick **evening* on my bed") when these errors were produced by a native speaker. When listening to non-native speech, semantic errors still elicited an N400 but syntactic violations from foreign-accented non-native speakers failed to elicit a P600 – a finding that was replicated for Chinese-accented English. The explanation that the authors adopt for these results is that listeners expect that non-native speakers will produce syntactic violations because of their lower second language proficiency, but do not expect non-native speakers to produce semantically bizarre utterances. Additional work using ERPs to investigate the effects of a foreign accent on semantic processing has found differences in the N400 component to native and non-native semantic errors, although results conflict as to whether the component was attenuated or amplified in non-native speech (Goslin, Duffy, & Floccia, 2012; Romero-Rivas, Martin, & Costa, 2015). Relatedly, very recent work shows that listeners tend to respond to semantically implausible sentences (e.g., "The mother gave the candle the daughter") as though they were their plausible counterparts ("The mother gave the candle to the daughter") when the sentences are uttered by non-native speakers (Gibson et al., 2017).

Non-native speaker status also has consequences about how the content of an utterance is evaluated offline. For example, non-native speakers are deemed less credible than their native speaker counterparts: general knowledge statements like "Ants don't sleep" that are true but not widely known are judged as less likely to be true when they are spoken by a non-native speaker with a thick foreign accent compared to when they are spoken by a native speaker (Lev-Ari & Keysar, 2010). Moreover, native listeners judge non-native speakers' narrative stories as more vague, are less likely to detect changes to a non-native speech stream in a change detection paradigm, and have poorer memory for sentences spoken by non-native speakers (Lev-Ari & Keysar, 2012). Even very young children hold negative biases towards non-native speakers – for example, they are less willing to befriend a non-native speaker and are more likely to trust a novel label offered by a native speaker as compared to one offered by a non-native speaker (Kinzler, Dupoux, & Spelke, 2007)

There are several possible explanations as to why these differences arise between non-native, foreign-accented speech and native speech (cf. Lev-Ari, 2015). One type of theory – which can be called the Intelligibility-Based account – is that a foreign accent is an additional processing demand that alters language comprehension because it is highly variable and perceptually distinct from the listener's own accent (e.g., Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Floccia, Goslin, Girard, & Konopczynski, 2006). A prediction that this account makes is that foreign-accented speech should be processed similarly to regional-accented speech and noisy speech. Intelligibility factors can explain the attenuated P600 (and intact N400, an earlier component that is arguably more automatic in nature than the P600 which reflects a reanalysis process; Kutas & Federmeier, 2011) to non-native speech observed by Hanulíková et al. (2012) as a result of cognitive overload, with few resources available for reanalysis of syntactic errors. As for offline effects, the Intelligibility-Based account would argue that non-native speakers are rated as more vague (Lev-Ari & Keysar, 2012) because they are actually more difficult to understand. Likewise, an Intelligibility-Based account could argue that participants in Gibson et al.'s (2017) study may have responded to semantically implausible sentences as though they were semantically plausible because of the costs imposed by processing a foreign accent.

Alternatively, what can be called Expectation-Based accounts argue that listeners have different expectations about the speech of non-native speakers from the outset; specifically, the expectation is that non-native speech is highly variable and that grammatical (and possibly semantic) errors will occur more often than in native speech (Lev-Ari, 2015; Niedzielski, 1999). These expectations cause individuals to rely more on top-down extra-linguistic information such as visual context and background knowledge of the situation. Expectation-Based accounts would argue that the results of Hanulíková et al. (2012) and Gibson et al. (2017) stem from the expectation that non-native speakers make more grammatical errors, leading to processing differences in the earliest moments of speech comprehension. Similarly, an Expectation-Based account would argue that listeners judge non-native speakers' narratives as more vague (Lev-Ari & Keysar, 2012) because of expectations held about the quality of non-native speech that affect speech processing.

Currently, researchers have yet to directly manipulate expectations about speaker identity (native vs. non-native) while keeping intelligibility constant. This type of manipulation would provide the strongest support for the role of expectations about speaker identity on the way language is processed. The difficulty in accomplishing this is that non-native speech, as has been noted above, is more challenging to understand compared to native speech. Thus in practice, the role of expectations about the speech of non-native speakers is difficult to isolate from the role of the processing cost incurred by a foreign accent. Here we resolve this difficulty by using a sentence rating task to compare how readers react to *written* sentences that they believe were uttered by a non-native vs. a native speaker. The advantage of using written materials is that processing demands arising from the sentences themselves are equivalent across the native and non-native speaker manipulations (across participants, the same sentences are attributed to different types of speaker). Thus, any asymmetries in how sentences are processed across speaker conditions can be unambiguously attributed to expectations about speaker identity. The present paradigm therefore allows us to isolate and probe the role of expectations on non-native speech processing in the absence of intelligibility factors.

1.1. Present study: pragmatic interpretation of native and non-native speech

Unlike previous work that has focused on syntactic or semantic processing of native vs. non-native speech (e.g., Hanulíková et al., 2012; Gibson et al., 2017), in the present study we focus specifically on how comprehenders interpret the pragmatic meaning of utterances produced by native vs. non-native speakers. Pragmatic aspects of meaning go beyond the semantic, literal meaning of a sentence and include contextual inferences that hearers compute as part of what the speaker intended to convey. Pragmatic aspects of meaning are driven by expectations about how rational communication works. Following Grice (1975), one can assume that interlocutors are mutually invested in a cooperative activity. According to Grice, listeners expect that their interlocutors aim to produce utterances that are true (Maxim of Quality), informative (Maxim of Quantity), relevant (Maxim of Relevance), and clear (Maxim of Manner). Because people strongly expect speakers to follow these maxims, they will often pragmatically enrich the literal semantic meaning of an utterance that appears to be in violation of the maxims, making an inference about what the speaker intended. For example, a sentence such as "Some giraffes have long necks" appears to violate the Maxim of Quantity: it is under-informative, because the speaker used the weaker term in a logical scale ('some') when s/he could have used a stronger, more informative scalar term ('all'). In many contexts, this utterance will lead the hearer to infer that not all giraffes have long necks (an inference known as *scalar implicature*; see Grice, 1975; Sperber & Wilson, 1986; Horn, 1972; Horn, 1984; Hirschberg, 1985; Carston, 1995; Levinson, 2000).

In the literature, judgments about under-informative sentences have been used as a test of whether a logical or pragmatic interpretation of the sentence has been reached: one might accept a sentence such as "Some giraffes have long necks" since the sentence is semantically/logically true; alternatively, one might reject the sentence since it pragmatically gives rise to a scalar implicature that is itself false ("Not all

giraffes have long necks"; Noveck, 2001). In general, judgment tasks that have used a 3-point or 5-point Likert scale have shown that adults (and even 5-year-old children) judge under-informative statements as more acceptable than completely false statements but not as good as completely true (and informative) statements (e.g., Katsos & Bishop, 2011; Davies, Andres-Roqueta, & Norbury, 2016). However, the degree to which comprehenders adopt logical or pragmatic interpretations varies with task demands and individual preferences (e.g., Bott & Noveck, 2004; Noveck, 2001; Guasti et al., 2005; Ozturk & Papafragou, 2016; Noveck & Posada, 2003; Feeney, Scrafton, Duckworth, & Handley, 2004; Hunt, Politzer-Ahles, Gibson, Minai, & Fiorentino, 2013; Tavano & Kaiser, 2010). It is unclear what individual characteristics contribute to this variability, but several options have been proposed, including social-communicative ability (Nieuwland, Ditman, & Kuperberg, 2010), executive function (De Neys & Schaeken, 2007), and participants' uncertainty about the Question under Discussion (Degen & Tanenhaus, 2015).

In this paper, we use a (non-binary) pragmatic judgment task to assess how expectations about non-native speakers affect comprehenders' interpretation of utterances produced by native and non-native speakers. Across three experiments, we present adult native speakers of English with written under-informative sentences and attribute these sentences to either native or non-native speakers of English. Participants then rate the sentences on the basis of how much sense they make. If altering beliefs (and corresponding expectations) about the speaker can change sentence interpretation, then judgements should change depending on speaker status. One possibility is that participants might judge under-informative (but true) sentences more negatively when uttered by non-native compared to native speakers for reasons related to biases against non-native speakers (cf. Lev-Ari & Keysar, 2010; Kinzler et al., 2007). Alternatively, under-informative statements might be given higher ratings when believed to have been produced by a non-native compared to a native speaker of English. Since non-native speakers are expected to be less accurate in their lexical (and other linguistic) choices, they may be seen as more likely to (unintentionally) produce under-informative utterances. Sins of information omission may thus be more likely to be forgiven in non-native speakers. This line of reasoning is in accordance with previous findings showing that listeners penalize grammatical violations less for non-native than for native speakers (Gibson et al., 2017; Hanulíková et al., 2012),

Because of the well-established variability in how people judge under-informative statements, we further investigate whether sensitivity to speaker identity in pragmatic judgments varies across the continuum of responding preferences (i.e., more logical vs. more pragmatic responders). One might expect that speaker sensitivity is higher in individuals who consistently respond to the pragmatically-enriched meaning of an utterance compared to those who tend to respond only to the literal meaning of an utterance within a task. This is because comprehenders who tend to adopt a pragmatic final interpretation recognize that the choice of one scalar term (e.g., 'some') over another (e.g., 'all') has pragmatic implications, and have reasoned about the alternatives that the speaker could have used but did not, as well as the reasons that the speaker must have had for using a less-than-optimal alternative (see Horn, 1972; 1984; Chierchia, Crain, Guasti, Gualmini, & Meroni, 2001; Gualmini, Crain, Meroni, Chierchia, & Guasti, 2001; Barner, Brooks, & Bale, 2011; Ozturk & Papafragou, 2015; Skordos & Papafragou, 2016). In the case of logical responders, alternatives to the present utterance may never have been considered (Bott & Noveck, 2004) or pragmatically-enriched meanings may have been considered, but later rejected in favor of a literal interpretation. Thus, more pragmatically-inclined responders may be more sensitive to properties of the speaker's identity and how these properties affect the choice of a linguistic stimulus and its intended meaning compared to people who tend to adopt a logical/semantic interpretation.

2. Experiment 1

In Experiment 1, we administered a Sentence Ratings task to compare how under-informative statements (among other types of statements) are processed when attributed to native vs. non-native speakers. We then investigated whether such speaker sensitivity varies across individuals. We also measured participants' general social-communicative ability and cultural attitudes towards non-native speakers, and related these measures to participants' ratings of pragmatic infelicities from different kinds of speakers.

2.1. Method

2.1.1. Participants

One hundred and fourteen native speakers of English aged 18–38 years (M = 28.14, SD = 4.16) living in the United States, 50 of whom were female, were recruited from Amazon's Mechanical Turk to participate in the experiment. Participants were compensated at a rate of $0.10 per minute for a total of $1.50.

2.1.2. Sentence ratings task materials

Eighty sentences were created for the Sentence Ratings task, half beginning with *some* and half beginning with *all*. Sentences were based on general knowledge and were evenly distributed across four Sentence Types: True but Under-Informative sentences with *some* (henceforth Under-Informative; "Some dogs are mammals"), True and Felicitous sentences with *some* (henceforth, True (Some); "Some people have dogs as pets in the house"), True and Felicitous sentences with *all* (henceforth, True (All); "All snow is cold and can melt into water"), and False sentences with *all* (henceforth, False; "All women are doctors who went to medical school"). The critical trials consisted of the Under-Informative sentences that were literally true but pragmatically odd (in the example above, all people have noses with two nostrils), and the other three Sentence Types were treated as control sentences. The four sentence types did not differ from one another in sentence length as measured in words or syllables (all p's $> .1$). All sentences used in this and all subsequent experiments are included in Supplemental Materials.

Speaker bios were created to accompany the sentences. Each bio either gave a short description of Emma, a native English speaker with a strong Boston accent (Native Speaker condition), or Yuqi, a native speaker of Mandarin Chinese with a strong Chinese accent (Non-Native Speaker condition). Thus, in both cases the speaker had an accent, and the only difference between the two was the non-native speaker status. There were two versions of each Speaker condition, in which the

Table 1
Speaker bios for Experiment 1.

Native Speaker	Non-Native Speaker
Emma is a college student at the University of Delaware, 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 Delaware from **Boston,** and her classmates often tease her about her **strong Boston accent**. She laughs it off, because she knows they are just having fun. In her spare time, **Emma** likes to **hike/run** and play the **piano/guitar**.	**Yuqi** is a college student at the University of Delaware, majoring in **History/Sociology**. She is doing well in her classes and plans to be a high school teacher after graduation. **Yuqi** moved with her family to Delaware from **China,** and her classmates often tease her about her **strong Chinese accent**. She laughs it off, because she knows they are just having fun. In her spare time, **Yuqi** likes to **hike/run** and play the **piano/guitar**.

speaker's hobbies and major varied. This was done so as not to present two nearly identical bios to the same participant. Thus there were four total bios, presented in Table 1.

Although the speaker bios for the Native and Non-Native Speaker conditions were nearly identical, we wanted to ensure that participants did not assume that one of the two speakers (or one of the two versions of the speakers) was more knowledgeable, particularly in terms of the topics in the critical Under-Informative sentences. Thus, we recruited an additional 60 participants from Mechanical Turk living in the United States. Participants read one of the four speaker bios and were then presented with each of the topics in the Under-Informative sentences (20 total). For example, "Dogs" would be the topic for the sentence "Some people have dogs as pets in the house." For each topic, participants rated on a scale from 0 to 100 how much they felt the person in the description knew about the topic. Mean ratings ($M = 61.42$, $SD = 9.38$) did not differ across the four speaker bios, nor did ratings for any one topic (all p's $> .1$). Thus, any potential differences between speakers in the Sentence Ratings task is unlikely to be attributed to perceptions of the speaker's general world knowledge.

2.1.3. Sentence ratings task procedure

The Sentence Ratings task consisted of two blocks: a Native Speaker block and a Non-Native Speaker block (counterbalanced across participants). Sentences within each block were evenly distributed across the four Sentence Types (10 of each), and were presented in a random order. Thus, both Speaker (Native, Non-Native) and Sentence Type (Under-Informative, True (Some), True (All), False) were treated as within-subjects factors. At the start of each block, one of the four speaker bios appeared on the screen. Participants were instructed to read carefully in order to answer the comprehension questions that followed, and were given as much time as they needed to read the paragraph before moving on. The speaker bio was followed by three multiple-choice questions about the speaker, presented in a random order ("Where is Emma/Yuqi from?", "What is Emma/Yuqi majoring in?", "What does Emma/Yuqi like to do in her spare time?"). Performance on these comprehension questions was quite high (88%), indicating that participants had fully read and understood the speaker bios. All participants correctly answered at least one of the two comprehension questions for each Speaker. Participants were then instructed that they would be reading 40 sentences that were originally uttered by the person they had just read about, and that their job was to rate how "Good" each sentence was on a five-point scale where 1 is "Very bad" and 5 is "Very good." Participants were instructed that a good sentence is one that makes perfect sense, and a bad sentence is one that makes no sense at all. Additionally, participants were told that because a given utterance can make more or less sense, they should make use of the intermediate values on the scale for sentences that were neither very good nor very bad.

On each trial, a sentence appeared in the center of the screen with the ratings scale below. The speaker bio was always present at the top of the screen, in a muted gray color. Participants could move the marker on the scale to indicate their desired rating. The marker snapped into one of five possible positions as it was moved (i.e., movement was not continuous). The five locations were not marked on the scale, but participants were instructed beforehand that there were five possible choices. As participants made their response, a face attached to the scale changed its expression (a frown for low ratings, a smile for high ratings, with three intermediate faces). Participants could take as long as they needed to make a response.

2.1.4. Autism-quotient questionnaire

Following the Sentence Ratings task, participants completed the Communicative Subscale of the Autism-Quotient Questionnaire (AQ-COMM; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). Originally designed with the aim of being a diagnostic tool for adolescents and adults on the Autism spectrum, the AQ-COMM has sometimes been shown to be correlated with scalar implicature computation (Nieuwland et al., 2010; Zhao, Liu, Chen, & Chen, 2015), even though the evidence is mixed (Antoniou, Cummins, & Katsos, 2016; Barbet & Thierry, 2016; Fairchild and Papafragou, 2018; Heyman & Schaeken, 2015). We included it in the present study under the hypothesis that the extent to which an individual rejects under-informative statements uttered by native and/or non-native speakers may be related to that individual's social communication skills, to the extent that they are measured by the AQ-COMM. The questionnaire consists of 10 statements (e.g., "I am often the last to understand the point of a joke," "I know how to tell if someone listening to me is getting bored"). For each statement, participants indicated how true it was of themselves. The standard scoring method was used, calculating a total score out of 10 of the number of autistic traits the person possessed.

2.1.5. Chinese cultural attitudes questionnaire

Finally, participants completed a Chinese Cultural Attitudes questionnaire, adapted from the American Attitudes Toward Chinese Americans & Asian Americans survey conducted by the Committee of 100. The questionnaire assesses how strongly individuals believe in cultural stereotypes of Chinese-Americans, both positive and negative. Participants were asked to rate how strongly they agreed with fourteen statements about Chinese-Americans (e.g., "Chinese-Americans are overly aggressive in the workplace," "Chinese-Americans have strong family values"). A total score was calculated for each participant based on the average agreement with cultural stereotypes.

2.2. Results

2.2.1. Overall analysis

Linear mixed-effects regressions were performed on Sentence Rating data for Experiment 1 and all subsequent experiments using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) for the R Project for Statistical Computing v3.2.2 (R Core Team, 2015). This method of analysis has several benefits, particularly for repeated-measures data like ours. For instance, variability across both participants and items can be accounted for in the same model, rather than needing to conduct separate by-participants and by-items analyses. For all this and all other mixed-effect analyses, predictor variables were sum-coded and p-values were obtained using likelihood ratio tests. Speaker (Native Speaker, Non-Native Speaker), Sentence Type (Under-Informative, True (Some), True (All), False), and the interaction between the two were included in the model as fixed effects, with crossed random intercepts for Participants and Items. Random slopes for Speaker and Sentence were also included in the model. Mean Sentence Ratings are presented in Fig. 1. Sentence Ratings differed significantly across Speakers, $\chi^2(1) = 13.303$, $p < .001$, and Sentence Types, $\chi^2(3) = 2517.640$, $p < .001$. Planned contrasts (presented in Table 2) indicate that Sentence Ratings were higher in the Non-Native Speaker ($M = 3.41$, $SD = 1.44$) condition as compared to the Native Speaker ($M = 3.32$, $SD = 1.46$) condition, $p < .001$. Additionally, Under-Informative ($M = 2.93$, $SD = 1.40$) sentences were rated higher than False ($M = 2.24$, $SD = 1.30$) sentences, $p < .001$, but lower than True (Some) ($M = 4.20$, $SD = 0.97$) sentences, $p < .001$. True (All) ($M = 4.09$, $SD = 1.05$) sentences were rated higher than True (Some) sentences, $p < .001$.

These main effects were qualified by a significant interaction between Speaker and Sentence Type, $\chi^2(3) = 16.849$, $p < .001$. Post-hoc tests (Bonferroni-corrected for multiple comparisons) indicated that ratings of Under-Informative sentences were higher in the Non-Native Speaker ($M = 3.02$, $SD = 1.39$) condition than in the Native Speaker ($M = 2.85$, $SD = 1.41$) condition ($p = .004$). In other words, participants were more accepting of under-informativeness when it was attributed to a Non-Native speaker. Ratings of True (Some), True (All) and False sentences did not differ by Speaker (all p's $> .05$), indicating that the effect was selective to under-informative sentences (and did not

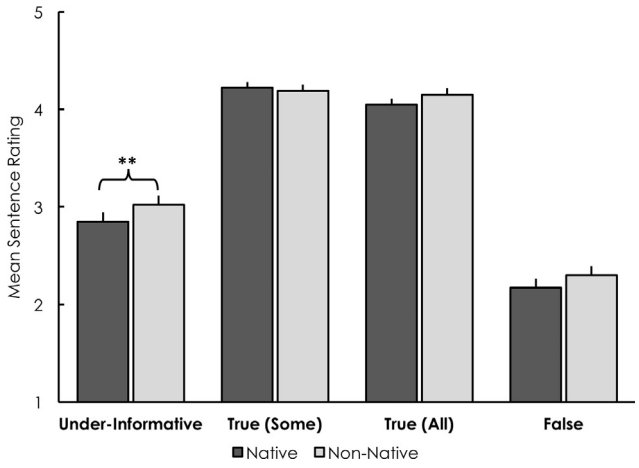

Fig. 1. Mean Sentence Ratings by Speaker for all Sentence Types in Experiment 1. Error bars indicate ± 1 S.E.M. Asterisks denote significance as follows: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table 2
Parameter estimates for a mixed-effects regression model predicting Sentence Ratings from Speaker and Sentence Type in Experiment 1.

Effect	β	*S.E.*	*t*	*p*
Intercept	3.490	0.036	97.990	< .001
Speaker (Native vs. Non-Native)	0.054	0.013	4.018	< .001
Sentence Type (Under-Informative vs. False)	1.147	0.058	19.811	< .001
Sentence Type (Under-Informative vs. True (Some))	−1.579	0.071	−22.313	< .001
Sentence Type (True (Some) vs. True (All))	−0.747	0.049	−15.227	< .001

extend to true or completely false statements).

2.2.2. Responder bias analysis

To further investigate the source of the forgiveness of under-informativeness in non-native speakers and its variation across individuals, a Non-Native Speaker Effect (hereafter NNS Effect) score was calculated for each participant by subtracting the mean rating for Under-Informative sentences in the Native Speaker condition from the mean rating for Under-Informative sentences in the Non-Native Speaker condition. Thus, individuals with positive scores were more lenient towards non-native speakers as compared to native speakers, while individuals with negative scores tended to penalize under-informativeness from non-native speakers more than from native speakers. To determine whether non-native speaker sensitivity varied across pragmatically- and logically-biased individuals, or whether the NNS Effect was stable across participants, we conducted a linear regression predicting the NNS Effect from the mean Under-Informative rating in the Native Speaker condition. This predictor was chosen because the Native Speaker condition reflects how a participant would judge under-informative utterances without other influences (and most closely corresponds to logical vs. pragmatic responders in the literature). As mentioned already, we expected that participants with a greater bias towards responding pragmatically (i.e., giving a low rating to under-informative sentences) in the Native Speaker condition may be more likely to take into account the speaker's identity (including the ability to handle linguistic alternatives) and be more accepting when a non-native speaker produces an under-informative utterance.

In our data, there was great variability in Under-Informative sentence ratings in the Native Speaker condition, with mean ratings ranging the entire span of the scale (1 – 5; *SD* = 1.01). Overall, as we had anticipated, sensitivity to under-informativeness in the Native Speaker

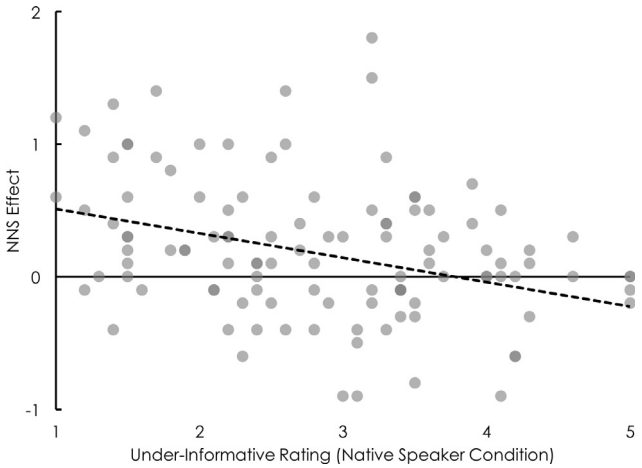

Fig. 2. NNS Effect by Under-Informative rating in the Native Speaker condition in Experiment 1. A NNS Effect of 0 indicates no difference in ratings between speaker conditions, whereas a positive NNS Effect indicates higher ratings of under-informativeness for Non-Native Speakers as compared to Native Speakers.

condition significantly predicted the NNS Effect, $F(1, 112) = 13.75$, $p < .001$, $R^2 = .11$. As can be seen in Fig. 2, the more a participant adopted a pragmatic interpretation of the under-informative utterances and judged them as not making sense when a native speaker uttered them, the more likely the participant was to give the benefit of the doubt to non-native speakers for such cases, $\beta = -0.184$, $SE = 0.050$, $t = -3.708$, $p < .001$.

2.2.3. Individual differences analyses

To further investigate potential individual differences in performance on the Sentence Ratings task, Kendall's tau correlation analyses (chosen to account for the positively-skewed distribution of AQ-COMM scores) were performed with the AQ-COMM score, the Chinese Cultural Attitude score, and Under-Informative Sentence Ratings for each Speaker as variables. AQ-COMM scores were marginally correlated with Under-Informative sentence ratings in the Native Speaker condition, $\tau_b(1\ 1\ 2) = .130$, $p = .058$, and not significantly correlated with such ratings in the Non-Native Speaker condition, $\tau_b(1\ 1\ 2) = .070$, $p > .1$. Thus, there was no evidence that Under-Informative sentence ratings are associated with social-communicative ability as measured by the AQ-COMM, in line with several previous results (Antoniou et al., 2016; Barbet & Thierry, 2016; Fairchild and Papafragou, 2018; Heyman & Schaeken, 2015). There was also no evidence that the differences in Under-Informative sentence ratings that we observed between speakers was due to an individual's cultural attitudes towards Chinese-American bilingual speakers: Chinese Cultural Attitude scores were not significantly correlated with Under-Informative sentence ratings for either the Native Speaker condition, $\tau_b(1\ 1\ 2) = -.036$, $p > .1$, or the Non-Native Speaker condition, $\tau_b(1\ 1\ 2) = .004$, $p > .1$.

2.3. Discussion

Three main findings arise from the present data. First, under-informative sentences were rated as making more sense than patently false sentences, but less sense than true sentences. Comprehenders thus understood that the under-informative statements were literally true, but were also sensitive to the fact that such statements were sub-optimal ways of conveying information. This finding constitutes a conceptual replication of nuanced pragmatic judgment patterns that have been previously observed for adults - and even children - in a laboratory setting (Katsos & Bishop, 2011).

Second, and critically, ratings of under-informative sentences increased when comprehenders believed these sentences to have come

from a non-native compared to a native speaker of English. This effect of speaker identity applied selectively to under-informative statements and did not extend to falsehoods (or simply true sentences), i.e., individuals did not overall give the benefit of the doubt to anything a non-native speaker said.

Third, participants who tended to consistently adopt a pragmatic interpretation of under-informative statements when uttered by a native speaker (and thus gave low ratings) were more forgiving towards non-native speakers for such sentences but those who tended to adopt the literal meaning for under-informative sentences (and thus gave high ratings) were less so. Assuming that calculating pragmatic inferences from the use of *some* requires reasoning about the communicative intentions of another person, including their access to linguistic alternatives such as *all*, it is reasonable to conclude that those individuals who consistently calculated the pragmatic meaning – unlike less pragmatically-inclined participants - were also sensitive to properties of the speaker (presumably reasoning, for instance, that a non-native speaker might not have been able to access or handle an alternative, pragmatically more felicitous way of phrasing their message).

The fact that comprehenders altered the way they processed under-informative statements simply as a result of information about speaker identity is in line with Expectation-Based accounts of non-native speech processing (predictions from Intelligibility-Based accounts do not speak to the present data since the sentences were presented in the written modality and there was no processing load difference between the Native and Non-Native Speaker condition). We hypothesize that the observed pragmatic lenience towards non-native speakers is related to comprehenders' beliefs about these speakers' linguistic competence. Suggestive, though not conclusive, evidence for this hypothesis comes from the fact that individuals who are likely to judge that under-informative sentences from native speakers make little sense (presumably because there are alternative, more felicitous means of constructing the sentence) also show the highest pragmatic lenience towards non-native speakers (presumably because these speakers lack the ability to handle such linguistic alternatives).

Several aspects of our findings argue against major alternative explanations of the speaker identity effect. For instance, since native and non-native speakers had been judged as equally knowledgeable of the subject matter in the under-informative sentences, it is unlikely that the effect of speaker identity could be attributed to differences in native vs. non-native speakers' general world knowledge (cf. also the lack of difference in tolerance for false statements attributed to native vs. non-native speakers). Furthermore, since there was no correlation between participants' social-communicative score or attitude towards Chinese individuals and their level of tolerance for pragmatic anomalies, neither general communicative skills nor cultural stereotypes appear to be likely sources of the pattern observed in our data. In the next experiment, we seek to strengthen and clarify the evidence linking forgiveness of non-native speakers' under-informativeness to those speakers' presumed L2 skills.

3. Experiment 2

In Experiment 2, we sought to replicate and extend evidence for the conclusion that comprehenders forgive under-informativeness to a greater extent from non-native as compared to native speakers. We followed the same procedure as in Experiment 1, but manipulated the degree to which the non-native speaker had an accent in English. In auditory studies, the strength of a non-native speaker's foreign accent is often interpreted as a marker of their second language proficiency (e.g., Kang, Rubin, & Pickering, 2010). If the results of Experiment 1 were due to expectations about the lower second language proficiency level of the non-native speaker, an accent-free speaker might be treated more closely to a native speaker compared to a non-native speaker with a heavy accent. An alternative possibility is that forgiveness of under-informativeness in non-native speakers emerges as a result of other, potentially cultural, attributions; if so, the pattern of results in our earlier experiment should extend to any kind of non-native speaker. The Chinese Cultural Attitudes Survey data in Experiment 1 suggest that cultural beliefs are unlikely to be the cause of participants' forgiveness of non-native speakers' under-informativeness, but in Experiment 2 we explicitly manipulate accent to provide a direct test of this idea.

3.1. Method

3.1.1. Participants

One hundred and eighty native speakers of English aged 20–35 (M = 29.33, SD = 3.91) living in the United States, 75 of whom were female, were recruited from Amazon's Mechanical Turk to participate in the experiment. Participants were compensated $2.00 for their time.

3.1.2. Materials

The materials were based on those in Experiment 1 with some minor alterations. First and foremost, we introduced an additional non-native speaker, Peiyao, who was also from China but had "no Chinese accent whatsoever." Thus, we had three within-subjects Speaker conditions: Native Speaker, Accent-Free Non-Native Speaker, and Accented Non-Native Speaker. For all three, we shortened the descriptions by removing the information about their performance in school and future career. There were three versions of each speaker bio, to add variation to the task, and as in Experiment 1 the majors and hobbies of the speaker were altered to create these different versions. All speaker bios for Experiment 2 are presented in Table 3.

An additional 40 sentences were created for the purposes of Experiment 2, 10 for each Sentence Type (Under-Informative, True (Some), True (All), False). These sentences followed the same constraints as in the previous experiment. Even with the addition of these sentences the four Sentence Types did not differ by length in words or syllables (all p's $>$ .1).

3.1.3. Procedure

We administered a Sentence Ratings task that was nearly identical to Experiment 1 except for the addition of a third block, for the Accent-Free Non-Native Speaker condition. As in the previous experiment,

Table 3
Speaker bios for Experiment 2.

Native Speaker	Accent-free Non-native speaker	Accented Non-native speaker
Emma is a college student at the University of Delaware, majoring in **History/Sociology/Mathematics**. **Emma** moved with her family to Delaware from **Boston**, and her classmates often tease her that she still has a **strong Boston accent**. In her spare time, **Emma** likes to **hike/run/swim** and play the **piano/guitar/violin**.	**Peiyao** is a college student at the University of Delaware, majoring in **History/Sociology/Mathematics**. **Peiyao** moved with her family to Delaware from **China** and her classmates often tease her about the fact that she has **no Chinese accent whatsoever.** In her spare time, **Peiyao** likes to **hike/run/swim** and play the **piano/guitar/violin**.	**Yuqi** is a college student at the University of Delaware, majoring in **History/Sociology/Mathematics**. **Yuqi** moved with her family to Delaware from **China**, and her classmates often tease her about her **strong Chinese accent**. In her spare time, **Yuqi** likes to **hike/run/swim** and play the **piano/guitar/violin**.

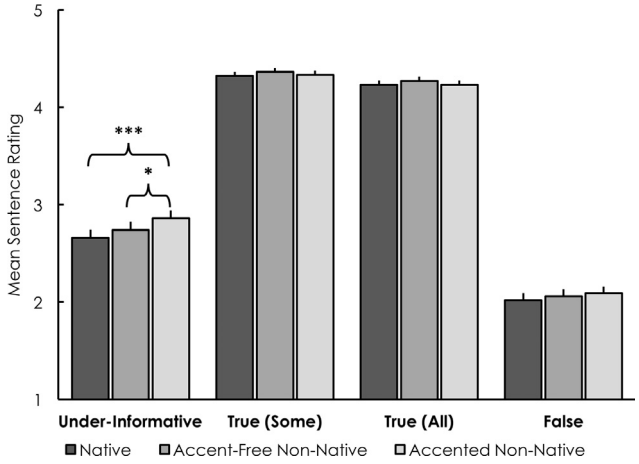

Fig. 3. Mean Sentence Ratings by Speaker for all Sentence Types in Experiment 2. Error bars indicate ± 1 S.E.M. Asterisks denote significance as follows: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

participants read a description of a speaker at the beginning of each block and answered comprehension questions about the speaker (performance was very high, 90%). Participants then judged 40 sentences "originally spoken by that person." They were asked to rate how "Good" each sentence was on a five-point scale where 1 is "Very bad" (makes no sense at all) and 5 is "Very good" (makes perfect sense). The order of the speaker was counterbalanced across participants, and sentences were fully rotated through each speaker condition.

3.2. Results

3.2.1. Overall analysis

A linear mixed-effects regression with crossed random intercepts for Participants and Items and Speaker (Native Speaker, Accent-Free Non-Native Speaker, Accented Non-Native Speaker) and Sentence Type (Under-Informative, True (Some), True (All), False) included as fixed effects with associated random slopes was performed on participants' Sentence Ratings in Experiment 2 (see Fig. 3). The effect of speaker was marginally significant, $\chi^2(1) = 5.422$, $p = .067$. Sentence Ratings varied significantly by Sentence Type, $\chi^2(3) = 7388.071$, $p < .001$. Planned contrasts (Table 4) indicated that ratings were higher in the Accented Non-Native Speaker ($M = 3.38$, $SD = 1.25$) condition as compared to the Native Speaker ($M = 3.31$, $SD = 1.30$) condition ($p = .031$), but ratings in the Accent-Free Non-Native Speaker ($M = 3.36$, $SD = 1.30$) condition did not differ significantly from the Native Speaker condition ($p > .1$). Under-Informative ($M = 2.76$, $SD = 1.11$) sentences were rated higher than False ($M = 2.06$, $SD = .94$) sentences, but lower than True (Some) ($M = 4.34$, $SD = .56$) sentences (both p's $< .001$). True (All) ($M = 4.25$, $SD = .57$) sentences were rated lower than True (Some) sentences ($p < .001$).

These main effects were qualified by a significant interaction between Speaker and Sentence Type, $\chi^2(3) = 28.053$, $p = .001$. Post-hoc tests revealed that Under-Informative sentences were rated significantly higher in the Accented Non-Native Speaker ($M = 2.86$, $SD = 1.49$) condition as compared to both the Native Speaker ($M = 2.66$, $SD = 1.49$) condition ($p < .001$) and the Accent-Free Non-Native Speaker ($M = 2.73$, $SD = 1.50$) condition ($p = .020$). Ratings did not differ significantly between the Native Speaker and Accent-Free Non-Native Speaker conditions ($p > .1$). In other words, participants treated non-native speakers with high linguistic competence like native speakers, and penalized their under-informative utterances. There was no difference in ratings for different kinds of Speakers in the other three Sentence types.

Table 4
Parameter estimates for a mixed-effects regression model predicting Sentence Ratings from Speaker and Sentence Type in Experiment 2.

Effect	β	S.E.	t	p
Intercept	3.349	0.034	97.333	< .001
Speaker (Native vs. Accented Non-Native)	−0.026	0.012	−2.167	.031
Speaker (Native vs. Accent-Free Non-Native)	−0.001	0.013	−0.126	.899
Sentence Type (Under-Informative vs. False)	1.189	0.059	20.211	< .001
Sentence Type (Under-Informative vs. True (Some))	−0.261	0.062	−4.208	< .001
Sentence Type (True (Some) vs. True (All))	0.590	0.61	9.627	< .001

3.2.2. Responder bias analysis

A NNS Effect score was calculated for each participant by subtracting mean Under-Informative Sentence Ratings in the Native Speaker condition from mean Under-Informative Sentence Ratings in the Accented Non-Native Speaker condition (the Accent-Free Non-Native Speaker condition was not included in calculation of the score, as these ratings did not differ significantly from the Native Speaker condition). A linear regression was performed predicting NNS Effect scores from Under-Informative Native Speaker Sentence Ratings. The analysis significantly predicted NNS Effect scores, $F(4, 178) = 19.55$, $p < .001$, $R^2 = .10$. As can be seen in Fig. 4, the more participants adopted a pragmatic interpretation of under-informative sentences (i.e., gave them a low rating), the more lenient they were towards the same under-informative statements when they were attributed to a non-native speaker with a strong accent, $\beta = -0.177$, $SE = 0.040$, $t = -4.421$, $p < .001$

3.3. Discussion

Experiment 2 replicated the general pattern of results in Experiment 1: regardless of whether they were attributed to a native or a non-native speaker, under-informative statements were judged as making more sense compared to completely false sentences, but less sense compared to true sentences. Additionally, Experiment 2 replicated the results of Experiment 1 by finding selectively higher ratings for under-informative statements believed to be produced by non-native compared to native speakers. Importantly, this effect was modulated by the language proficiency of the non-native speaker: under-informative sentences were judged as making more sense when they came from an accented non-native speaker compared to a native speaker but an

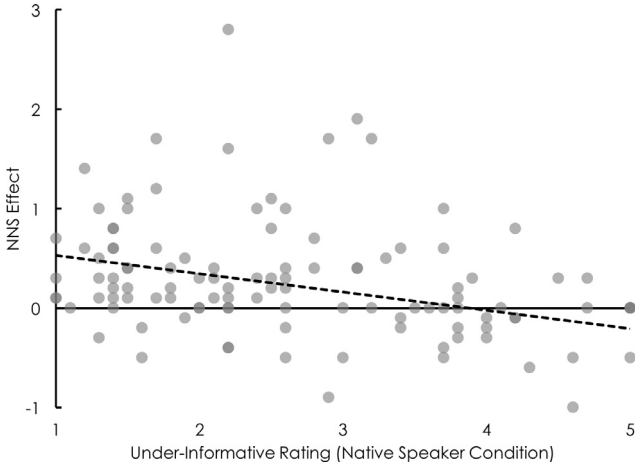

Fig. 4. NNS Effect by Under-Informative rating in the Native Speaker condition in Experiment 2. A NNS Effect of 0 indicates no difference in ratings between speaker conditions, whereas a positive NNS Effect indicates greater lenience towards under-informativeness from Accented Non-Native Speakers as compared to Native Speakers.

accent-free non-native speaker had no such advantage. Finally, we replicated the finding that this selective advantage for non-native speakers was greater in those participants who consistently derived pragmatic inferences from under-informative statements.

As with Experiment 1, our findings strongly support an Expectation-Based account of non-native speech processing: different responses to the same under-informative sentences across conditions were produced simply by altering beliefs about the language background of the speaker. Furthermore, the present data suggest that it is expectations about the language proficiency of the speaker specifically (and not, e.g., cultural attitudes) that lead to greater forgiveness of under-informativeness. We hypothesize that, given accent information alone, comprehenders make further assumptions about the non-native speakers' L2 proficiency level (cf. Kang et al., 2010) and go on to assume that only non-native speakers with a poor command of their second language should be given the benefit of the doubt when they produce under-informative statements.

4. Experiment 3

The under-informative sentences used in Experiments 1 and 2 (e.g., "Some people have noses with two nostrils") relied on world knowledge (e.g., knowing that noses have two nostrils). Furthermore, the corresponding "not all" propositions ("Not all people have noses with two nostrils") were false and unlikely to be part of speaker meaning. In Experiment 3, we sought to replicate the findings of Experiments 1 and 2 using a different set of stimuli for which judgments did not rely on evaluating individual sentences against one's own world knowledge; furthermore, the "not all" propositions could plausibly have been intended to be part of what the speaker meant by uttering *some*. Building on materials used in a study by Bergen and Grodner (2012), we introduced three-sentence passages where a highly knowledgeable speaker used *some* in a way that was highly likely to give rise to the "not all" implicature. The "not all" implicature was either cancelled explicitly ("In fact, all…") or supported ("The rest…") in the final sentence of the passage.

In Bergen and Grodner's (2012) reading study (whose paradigm was based on Breheny, Katsos, & Williams, 2006), speaker knowledge was manipulated in a context sentence and participants showed sensitivity to speaker knowledge that manifested itself in their reading times. Participants generated stronger implicatures when the speaker was highly knowledgeable of the topic at hand ("I *meticulously compiled* the investment reports. Some of the real estate investments lost money.") and therefore were more likely to have meant that the stronger alternative *all* was false. Conversely, participants generated weaker implicatures in cases when the speaker was less knowledgeable ("I *skimmed* the investment reports. Some of the real estate investments lost money.") and therefore may not have meant that the stronger alternative was false, but simply implicated lack of knowledge about whether the stronger alternative was false. In the present experiment, we only used cases where the speaker was highly knowledgeable to focus strictly on the effects of language background: the passages were preceded by information about the speaker (native vs. non-native) and participants were asked to rate the passages for meaning, as in our prior studies. We reasoned that participants would take a non-native speaker to be more likely compared to a native speaker to make (and later correct) a pragmatically under-informative statement, presumably because of poorer initial choice of words due to lower language proficiency/pragmatic competence.

4.1. Method

4.1.1. Participants

One hundred and ten English monolinguals aged 20–42 years (M = 28.62, SD = 4.34), living in the United States were recruited from Amazon's Mechanical Turk (n = 46 Female, n = 62 Male, n = 2 Other/Prefer not to answer). Participants were compensated $1.50 for the fifteen-minute study.

Table 5
Sample stimuli for Experiment 3.

Passage Type	Example
Some/All	As part of my advanced accounting class, I meticulously compiled the investment reports. **Some** of the investments lost money. **In fact, they all did** because of the recent economic downturn.
Some/Rest	As part of my advanced accounting class, I meticulously compiled the investment reports. **Some** of the investments lost money. **The rest** did not despite the recent economic downturn.
Only some/Rest	As part of my advanced accounting class, I meticulously compiled the investment reports. **Only some** of the investments lost money. **The rest** did not despite the recent economic downturn.
Only some/All	As part of my advanced accounting class, I meticulously compiled the investment reports. **Only some** of the investments lost money. **In fact, they all did** because of the recent economic downturn.

4.1.2. Materials and procedure

Forty passages were created for Experiment 3, each with 4 versions (Some/All, Some/Rest, Only some/Rest, Only some/All; see Table 5). Twenty-two passages came from the stimulus list in Bergen and Grodner (2012), with minor word changes. The rest were created for the experiment and followed the same structure as the passages borrowed from Bergen and Grodner (2012). Passages were created such that it would be believable for a college student to have produced them. All passages began with a context sentence establishing that the speaker was fully knowledgeable about the topic (e.g., "As part of my advanced accounting class, I meticulously compiled the investment reports." – see Table 5.)

In Some/All passages, the context sentence was followed by a critical sentence beginning with *some* meant to trigger a "not all" implicature but the final sentence cancelled the implicature ("In fact… all.") Some/Rest passages included the same critical sentence as the Some/All passages but their final sentence was consistent with the implicature ("The rest…"). The Only some/Rest passages were identical to the Some/Rest passages, except that the critical sentence began with *Only some* instead of *some*, and therefore, there was no need to calculate an implicature. The Only some/All passages were identical to the Some/All passages, except that again the critical sentence began with *Only some* instead of *some*. In this case, not only was there no implicature to be generated but the final sentence beginning with "In fact…all" was logically inconsistent with the critical sentence ("Only some…").

The task consisted of two blocks: a Native Speaker block and a Non-Native Speaker block (counterbalanced across participants), using the same speaker descriptions as in Experiment 1. Passages within each block were evenly distributed across the four conditions (10 of each), and were presented in a random order. Thus, both Speaker (Native, Non-Native) and Passage Type (Some/All, Some/Rest, Only some/Rest, Only some/All) were treated as within-subjects factors. At the start of each block, one of the four speaker bios appeared on the screen. Participants were instructed to read carefully in order to answer the comprehension questions that followed, and were given as much time as they needed to read the paragraph before moving on. The speaker bio was followed by three multiple-choice questions about the speaker, presented in a random order ("Where is Emma/Yuqi from?", "What is Emma/Yuqi majoring in?", "What does Emma/Yuqi like to do in her spare time?"). Performance on these comprehension questions was very good (87%), indicating that participants had fully read and understood the speaker bios. Participants were then instructed that they would be reading 40 passages that were originally uttered by the person they had just read about, and that their job was to rate how "Good" each sentence was on a five-point scale where 1 is "Very bad" (makes no sense at

all) and 5 is "Very good" (makes perfect sense). On each trial, a three-sentence passage appeared in the center of the screen with the ratings scale below. For half of the participants, the speaker bio was always present at the top of the screen in a muted gray color, and for the other half it was not present at the top of the screen. Results are combined for these two groups as they did not differ from one another in terms of sentence ratings for any condition.[2]

We predicted that Some/Rest and Only some/Rest passages would elicit high ratings, as both types of passages make sense and are pragmatically felicitous (cf. the True sentences in our previous experiments). Only some/All passages should elicit the lowest ratings, as the final sentence logically contradicts and corrects the critical sentence (cf. our earlier False sentences). Some/All passages should elicit higher ratings than Only some/All passages but lower ratings than the other two conditions because of the presence of an under-informative statement that is later corrected (cf. the Under-Informative sentences in our previous experiments).

For our critical Speaker manipulation, we expected participants to rate Some/All passages more highly in the Non-Native Speaker condition compared to the Native Speaker condition. This would indicate that participants would be more accepting of a Non-Native speaker inadvertently producing an utterance in which *some* is compatible with *all*, mirroring the results of Experiments 1 and 2. No such differences were expected in the other passages.

4.2. Results

4.2.1. Overall analysis

A linear mixed-effects regression with crossed random intercepts for Participants and Items and Speaker (Native Speaker, Accent-Free Non-Native Speaker, Accented Non-Native Speaker) and Sentence Type (Some/All, Some/Rest, Only some/Rest, Only some/All) included as fixed effects with associated random slopes was performed on participants' Sentence Ratings in Experiment 3 (see Fig. 5). Sentence Ratings varied significantly by Speaker, $\chi^2(1) = 6.696$, $p = .010$, and by Sentence Type, $\chi^2(3) = 2448.791$, $p < .001$. Planned contrasts (Table 6) indicated that ratings were higher in the Non-Native Speaker ($M = 3.38$, $SD = 1.24$) condition as compared to the Native Speaker ($M = 3.26$, $SD = 1.22$) condition ($p = .008$). Furthermore, Some/All ($M = 2.76$, $SD = 1.80$) passages were rated higher than Only some/All ($M = 2.40$, $SD = 1.31$) passages but lower than Some/Rest ($M = 4.02$, $SD = 1.17$) passages (both p's $< .001$). Only some/Rest ($M = 4.05$, $SD = 1.23$) passages were rated higher than Some/Rest passages ($p < .001$).

These main effects were qualified by a significant interaction between Speaker and Passage Type, $\chi^2(3) = 40.823$, $p < .001$. Post-hoc tests indicated that ratings of Some/All passages were higher for Non-Native speaker trials ($M = 2.98$, $SD = 1.83$) than for Native speaker ($M = 2.53$, $SD = 1.30$) trials ($p < .001$). Ratings of Some/Rest, Only some/Rest, and Only some/All passages did not differ by Speaker (all p's $> .1$).

4.2.2. Responder bias analysis

A NNS Effect score was calculated for each participant by subtracting mean Some/All ratings in the Native Speaker condition from mean Some/All ratings in the Non-Native Speaker condition. A linear regression was then performed predicting NNS Effect scores from mean Some/All ratings in the Native Speaker condition, as in the previous experiments: the analysis yielded a significant result, $F(1, 108) = 18.03$, $p < .001$, $R^2 = .14$. As Fig. 6 demonstrates, the worse an individual rated a passage in which *some* was compatible with *all* (i.e., made a pragmatic judgment rather than a logical one), the more lenient they were towards such passages when produced by a non-native speaker, $\beta = -0.251$, $SE = 0.059$, $t = -4.246$, $p < .001$.

[2] A mixed ANOVA was conducted with Speaker (Native, Non-Native) and Sentence Type (Some/All, Some/Rest, Only some/All, Only some/Rest) as within-subjects factors and Experiment (Paragraph Present on every trial or Absent) as a between-subjects factor. The main effect of Experiment was not significant, nor were any interactions with Experiment (all p's $> .05$).

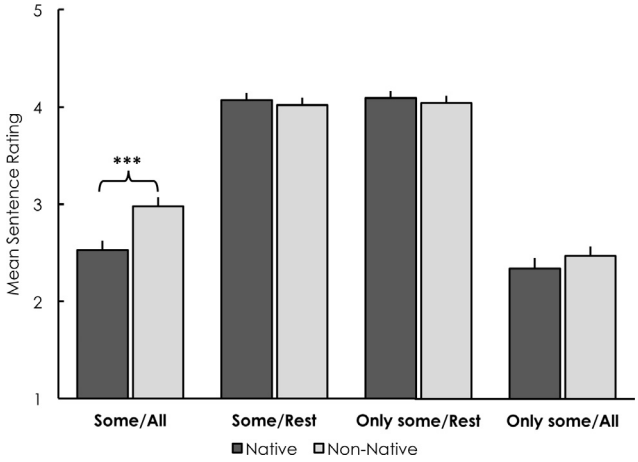

Fig. 5. Mean Sentence Ratings by Speaker for all Passage Types in Experiment 3. Error bars indicate ± 1 S.E.M. Asterisks denote significance as follows: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Table 6
Parameter estimates for a mixed-effects regression model predicting Sentence Ratings from Speaker and Sentence Type in Experiment 3.

Effect	β	S.E.	t	p
Intercept	3.259	0.039	86.683	< .001
Speaker (Native vs. Non-Native)	0.052	0.019	2.710	.008
Sentence Type (Some/All vs. Only Some/All)	1.001	0.053	18.777	< .001
Sentence Type (Some/All vs. Some/Rest)	−1.634	0.077	−21.102	< .001
Sentence Type (Some/Rest vs. Only Some/Rest)	−0.835	0.044	−19.118	< .001

4.3. Discussion

In line with the pattern of results previously observed, under-informative passages with *some* followed by *all* were treated as making more sense than logically inconsistent passages in which *only some* was followed by *all*, but less sense than passages in which *some* or *only some* was followed by *the rest*. Importantly, comprehenders judged that these under-informative passages made more sense when they believed them to have come from a non-native compared to a native speaker of English. As in Experiments 1 and 2, this effect of speaker identity was selective to Some/All passages and did not extend to any of the other conditions, and was greatest in participants who tended to give lower meaning ratings to the Some/All statements (i.e., participants who had adopted a pragmatic interpretation of *some* upon first reading it). Experiment 3 therefore fully replicated our earlier findings with a new set of stimuli, showing that forgiveness of non-native speakers' under-informativeness is robust, generalizable, and does not rely on world knowledge.

5. General discussion

5.1. Theories of non-native language processing

Interacting with non-native speakers poses specific challenges for language processing, with prior research indicating that neural and behavioral responses to native and non-native language errors differ

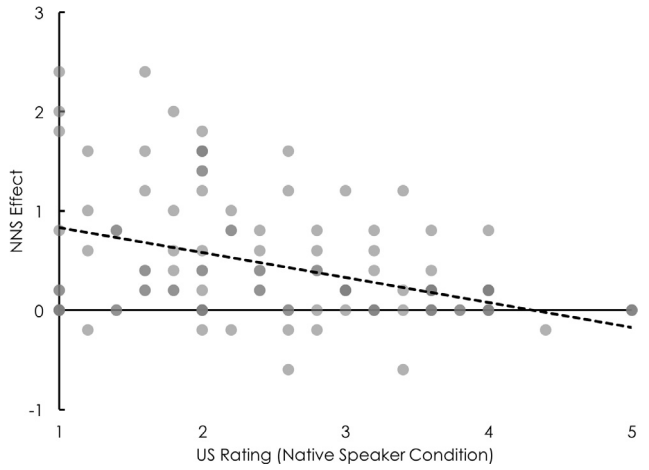

Fig. 6. NNS Effect by Some/All rating in the Native Speaker condition in Experiment 3. A NNS Effect of 0 indicates no difference in ratings between speaker conditions, whereas a positive NNS Effect indicates greater tolerance of under-informativeness from Accented Non-Native Speakers as compared to Native Speakers.

(Gibson et al., 2017; Goslin et al., 2012; Hanulíková et al., 2012; Romero-Rivas et al., 2015), and that listeners judge non-native speakers to be less trustworthy and more vague (Lev-Ari & Keysar, 2012). Intelligibility-Based accounts argue that non-native speech is processed differently to the extent that it requires more processing resources. Expectation-Based theories argue that differences in online and offline processing of non-native speech stem from the different expectations that listeners hold about non-native speakers (e.g., that non-native speakers have lower language proficiency, or that their speech stream will be more variable). In practice, the role of expectations is hard to disentangle from the intelligibility costs incurred by a foreign accent. Here we presented a strong test of the role of expectations in the absence of any actual intelligibility-related costs by comparing ratings of sentences presented in the written modality but believed to have been produced by either a native or a non-native speaker. Furthermore, we broadened the empirical scope of prior work on the comprehension of non-native speech by focusing on the domain of pragmatics.

Across three experiments, we found that knowledge about the language background of the speaker affected pragmatic interpretation even in the absence of actual exposure to a foreign accent. Specifically, comprehenders rated pragmatically under-informative sentences (e.g., “Some people have noses with two nostrils”) as making more sense when they believed that these sentences were produced by a non-native speaker as compared to a native speaker (Experiment 1). The effect was present for non-native speakers with lower second language proficiency but not for highly proficient (non-accented) non-native speakers (Experiment 2). Furthermore, the native vs. non-native speaker difference extended to an additional set of stimuli that did not rely on world knowledge (Experiment 3). Throughout these experiments, speaker sensitivity was related to individual judgment preferences: individuals who tended to respond to the pragmatic meaning of under-informative statements when these statements were attributed to a native speaker (and hence based their judgment on the presence of a more informative alternative) were most forgiving of non-native speakers’ under-informativeness - presumably because these speakers’ access or ability to evaluate alternatives was impaired; by contrast, individuals who tended to respond to the logical meaning of under-informative statements when statements belonged to a native speaker (and hence did not focus on the presence of better linguistic alternatives) were less likely to adjust their ratings for non-native speakers. Together, our findings strongly support Expectation-Based accounts of non-native language processing, even in the absence of intelligibility factors. Our results are in line with previous studies that suggest a role for speaker identity expectations in non-native speech processing (Goslin et al., 2012; Hanulíková et al., 2012; Lev-Ari, 2015; Romero-Rivas et al., 2015).

Naturally, the present data are entirely compatible with the idea that a foreign accent introduces noise to the linguistic signal and increases processing effort. In spoken communication, non-native input to language comprehension is often an errorful or corrupted signal on multiple levels, and it as such can incur intelligibility-based cost. Communicating with non-native speakers typically makes use of both general expectations about the language and error patterns of non-native speakers of the kind discussed here, as well as situation-specific experiences with and adaptations to actual error patterns in the speech of the particular individual one is communicating with (on such situation-specific factors, see Gibson et al., 2017).

5.2. The pragmatics of accent

How can the increased tendency to forgive sins of information omission in non-native speakers be reconciled with the finding that non-native speakers’ utterances are often judged to be less trustworthy and more vague compared to those of native speakers (e.g., Lev-Ari & Keysar, 2010)? One might expect an under-informative utterance coming from a non-native speaker to be judged less charitably or corrected more often by native comprehenders. We want to note that there have been cases where the linguistic instability inherent in much of non-native speech has been found to have some advantages: as mentioned already, syntactically errorful utterances are less likely to elicit surprise (Hanulíková et al., 2012) and more likely to be reinterpreted when produced by non-native speakers (Gibson et al., 2017). For our data, we believe that non-native speakers are penalized less because they (are perceived to) have reasonable grounds for selecting a less-than-optimal linguistic stimulus – namely, they are linguistically less competent.

The computations leading to the pragmatic lenience effect are worth discussing in some detail. Recall that pragmatic aspects of meaning are driven by expectations about how rational communication works. For instance, communicators expect speakers to strive to offer sentences that are informative to the degree required by the goals of the conversation (Maxim of Quantity; Grice, 1975). When speakers fail to be as informative as required, hearers are justified to engage in further inferences to understand the reasons behind this failure. In some cases, listeners derive a scalar implicature, inferring that the speaker meant that a more informative statement would not be true. In other cases, under-informative statements give rise to the inference that the speaker was unable to commit to the stronger term because of lack of information (see also Bonnefon, Feeney, & Villejoubert, 2009; Geurts, 2010; Sperber & Wilson, 1995, for additional possibilities).

In Experiments 1 and 2, statements such as “Some humans have noses with two nostrils” violate the Maxim of Quantity: the speaker used the weaker term in a logical scale (‘some’) when she could have used a stronger, more informative scalar term (‘all’). Furthermore, the statements can be potentially misleading because they can give rise to a scalar inference corresponding to a patently false proposition (“Not all humans have noses with two nostrils”). After presumably detecting the violation, comprehenders judge under-informative sentences as “making less sense” when attributed to a native speaker because it is hard to perceive what the speaker could have meant (i.e., what the grounds for under-informativeness could be given what is known about Emma’s abilities and preferences). For a non-native speaker, comprehenders’ judgments are more charitable (even though not completely positive) since the infelicitous scalar choice could be attributed to lack of proficiency in English. (The same considerations apply to Experiment 3: here the speaker explicitly corrects ‘some’ to ‘all’ precisely because the earlier choice of quantifier was likely to lead to a misleading inference.) Comprehenders are therefore more likely to forgive non-native speakers for sins of information omission (for these speakers know not what they do.) As our data consistently show, the tendency to

forgive is stronger in participants who tend to adopt a pragmatic final interpretation – probably because these participants base their rating on the alternatives that the speaker could have used but did not. Notice that these comprehenders compute the pragmatic content derivable from the speaker's utterance ("Not all…") even though they are unlikely to believe it themselves – and may not assume that it was meant to be communicated by the speaker (cf. Mazzarella, 2015; Sperber et al., 2010).[3]

It is worth noting that non-native speaker identity does not completely override pragmatic interpretation, rather it results in a *shift* in judgments: even those comprehenders who were more lenient towards non-native speakers' under-informativeness did not find such statements to be as acceptable as true statements. Interestingly, as can be seen in Figs. 2, 4, and 6, some participants even rated under-informative statements uttered by non-native speakers lower than under-informative statements from native speakers. Although we did not find a relationship between cultural attitudes and sentence ratings, an intriguing possibility is that the comparatively lower ratings that some participants gave to non-native speakers' utterances stemmed from some type of negative bias towards non-native speakers.

The present evidence for pragmatic lenience towards non-native speakers comes from an offline judgment task. Such tasks have proven very useful as a means of investigating pragmatic intuitions in both adults and young children (see Katsos & Bishop, 2011). Nevertheless, we anticipate that comprehenders beyond the present context make spontaneous assumptions about why people are less informative than expected and, as part of these computations, attribute different grounds for under-informativeness to native and non-native speakers in a variety of tasks. Our ongoing work currently supports this prediction (Fairchild & Papafragou, 2018).

5.3. Extensions and future directions

Our data suggest several possibilities for future research. First, all of our experiments compared native speakers of English to native speakers of Chinese who spoke English as a second language. It remains open whether (descriptions of) different types of accents are equally likely to induce adjustments in pragmatic processing. Relatedly, it remains to be seen whether such adjustments emerge regardless of the comprehenders' specific language background (if so, the reported effect would be replicated in reverse with participants recruited in China). Versions of the present experiments could pursue these questions by varying both the language of the comprehenders and the language background (and level of proficiency) of the presumed non-native speakers. From a broader perspective, it is intriguing to explore whether selective pragmatic lenience of the kind discussed here would generalize to tokens produced by other populations whose linguistic knowledge or use is developing, atypical or otherwise limited (examples include children acquiring their first language, or aphasic patients).

Second, effects of (non)native speaker status in our data were observed selectively in under-informative sentences but not in true (and informative) or false sentences. We suspect that this selectivity is due to the fact that non-native speakers in our studies were introduced as highly educated, already living abroad and being members of a university community; furthermore, the experimental sentences were fairly sophisticated (especially in Experiment 3) and contained no grammatical errors. Any differences between the two groups of speakers was therefore limited to relatively nuanced aspects of communication. This conclusion is reinforced by the fact that the two groups of speakers did not differ in their perceived (Experiment 1) or stated (Experiment 3) familiarity with the sentence topics so there was no basis for assuming that they differed in their ability to judge a test sentence as factually true or not. It remains possible that, in populations of non-native speakers with less secure knowledge of the mechanics of their second language, the observed lenience might extend to semantic errors (see Goslin et al., 2012).

Third, the present results cohere with a broader perspective according to which accents are not just physical features of a linguistic stimulus but sources of psychological attributions. Accents can form the basis of assumptions about the speaker's epistemic state, cultural beliefs, experience with food, music and the environment, and several other attributes beyond language. Depending on the topic, accented speakers may be considered more, not less knowledgeable than native speakers and these epistemic assumptions can themselves bear on utterance interpretation. For instance, if a Chinese-accented person used a scalar utterance in discussing Chinese politics (e.g., "Some Chinese families follow the one-child policy"), the listener's comprehension of her utterance would probably be affected by her presumed expertise (for instance, the listener may conclude that the speaker meant that not all families follow the policy, whereas the same utterance from a native speaker of English might be taken to convey lack of knowledge about the situation in all families in China). In the present work, knowledge of the topics in test sentences was comparable across native and non-native speakers (see, e.g., Experiments 1 and 3). Future work could fruitfully investigate how situational knowledge or cultural attitudes interact with the non-native speaker effect we observed here.

5.4. Final thoughts

Viewed most broadly, our findings contribute to a long line of evidence demonstrating that both speaker and listener identity strongly modulate language processing (e.g., Just & Carpenter, 1992; Prat, Keller, & Just, 2007; van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008; Nakano, Saron, & Swaab, 2010; Regel, Coulson, & Gunter, 2010; Prat & Just, 2011; Boudewyn, Long, & Swaab, 2012; Kamide, 2012; Tanner & Van Hell, 2014). For example, van Berkum et al. (2008) found that speaker identity affected online semantic integration: well-formed sentences such as "Every evening I drink wine before I go to bed" with no apparent semantic violations led to increased N400 responses when produced by an unlikely speaker given one's world knowledge (e.g., a young child). Later related work has shown that the ability to integrate such world knowledge during sentence processing varies with a listener's cognitive abilities, such as working memory capacity (e.g., Nakano et al., 2010). While many models of language comprehension take into account individual variation in listener characteristics (Gernsbacher, Varner, & Faust, 1990; Just & Carpenter, 1992), the strong evidence for speaker sensitivity reported here supports models which can also account for the use of speaker properties in language processing (e.g., Nadig & Sedivy, 2002; Sedivy, 2007; Brennan & Hanna, 2009; Pickering & Garrod, 2013).

Our data leave open whether non-native speaker status is integrated on-line during the earliest stages of sentence processing to guide

[3] The present perspective differs from (and is orthogonal to) the notion of pragmatic tolerance developed by Katsos and Bishop (2011) to account for the fact that, unlike adults who penalize under-informative utterances, young children seem to find them acceptable when given a binary response scale (when given a 3-point scale, the difference disappears, and both age groups give under-informative utterances intermediate ratings). According to Katsos and Bishop, young children detect under-informativeness in binary tasks but – unlike adults - do not deem it serious enough to warrant a negative judgment. Thus in their account, the notion of tolerance is meant to explain task-specific behavior, i.e., children's apparently logical responses to under-informativeness *within a binary judgment task*. In the present data, lenience towards non-native speakers leads to *higher*, not lower, ratings for under-informativeness in adults. Crucially, these higher ratings are not taken to reflect task-specific reasoning (or a difference in how task demands are interpreted in the Native vs. Non-Native speaker conditions) but rather specific inferences about the grounds of under-informativess in speakers of different linguistic backgrounds. Finally, and relatedly, such inferences are not meant to be limited to metalinguistic contexts but should arise spontaneously when people process non-native speech that falls short of informativeness expectations (see also Fairchild & Papafragou, in prep.

pragmatic inference or affects later stages of processing. There is evidence that other properties of the speaker such as speaker knowledge of the situation at hand are integrated early during sentence processing (Bergen & Grodner, 2012; Breheny, Ferguson, & Katsos, 2013). Electrophysiological studies indicate that the non-native status of the speaker affects syntactic processing online (Goslin et al., 2012; Hanulíková et al., 2012; Romero-Rivas et al., 2015; Grey & Van Hell, 2016), but it remains to be seen whether the same is true in the domain of pragmatics.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.cognition.2018.08.010.

References

Antoniou, K., Cummins, C., & Katsos, N. (2016). Why only some adults reject underinformative utterances. *Journal of Pragmatics, 99*, 78–95.

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America, 133*(3), EL174–EL180.

Barbet, C., & Thierry, G. (2016). Some alternatives? Event-related potential investigation of literal and pragmatic interpretations of some presented in isolation. *Frontiers in Psychology, 7*, 1479.

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition, 118*(1), 84–93.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5–17.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America, 114*(3), 1600–1610.

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.

Bialystok, E., Craik, F. I., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*(2), 290–303.

Bonnefon, J. F., Feeney, A., & Villejoubert, G. (2009). When some is actually all: Scalar inferences in face-threatening contexts. *Cognition, 112*(2), 249–258.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language, 51*(3), 437–457.

Boudewyn, M. A., Long, D. L., & Swaab, T. Y. (2012). Cognitive control influences the use of meaning relations during spoken sentence comprehension. *Neuropsychologia, 50*(11), 2659–2668.

Breheny, R., Ferguson, H. J., & Katsos, N. (2013). Taking the epistemic step: Toward a model of on-line access to conversational implicatures. *Cognition, 126*(3), 423–440.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition, 100*(3), 434–463.

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science, 1*(2), 274–291.

Carston, R. (1995). Quantity maxims and generalized implicature. *Lingua, 96*, 213–244.

Census Bureau, U. S. (2011). *American community survey reports.* Washington, DC: U.S. Census Bureau.

Chierchia, G., Crain, S., Guasti, M. T., Gualmini, A., & Meroni, L. (2001). *The acquisition of disjunction: Evidence for a grammatical view of scalar implicatures* (pp. 157–168). Somerville, MA: Cascadilla Press.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America, 116*(6), 3647–3658.

Clyne, M. (1987). Constraints on code switching: How universal are they? *Linguistics, 25*(290), 739–764.

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition, 113*(2), 135–149.

Davies, C., Andrés-Roqueta, C., & Norbury, C. F. (2016). Referring expressions and structural language abilities in children with specific language impairment: A pragmatic tolerance account. *Journal of Experimental Child Psychology, 144*, 98–113.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General, 134*(2), 222–241.

De Bot, K., & Schreuder, R. (1993). *The bilingual lexicon* (pp. 191–214). .

De Neys, W., & Schaeken, W. (2007). When people are more logical under cognitive load: Dual task impact on scalar implicature. *Experimental Psychology, 54*(2), 128–133.

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature A constraint-based approach. *Cognitive Science, 39*(4), 667–710.

Fairchild, S., & Papafragou, A. (2018). Why didn't you say so? Information omission in native and non-native speakers. Ms., University of Delaware (in preparation).

Fairchild, S., & Papafragou, A. (2018). The roles of executive function and theory of mind in pragmatic computations (submitted for publication).

Feeney, A., Scrafton, S., Duckworth, A., & Handley, S. J. (2004). The story of some: Everyday pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology, 58*(2), 121.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41*(1), 78–104.

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance, 32*(5), 1276–1293.

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 430–445.

Geurts, B. (2010). *Quantity implicatures.* New York: Cambridge University Press.

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.

Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain and Language, 109*(2), 55–67.

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language, 122*(2), 92–102.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(02), 67–81.

Grice, H. P. (1975). *Syntax and semantics* (pp. 41–58). .

Grosjean, F. (2010). *Bilingual: life and reality.* Cambridge, MA: Harvard University Press.

Gualmini, A., Crain, S., Meroni, L., Chierchia, G., & Guasti, M. T. (2001, October). At the semantics/pragmatics interface in child language. In Semantics and Linguistic Theory (Vol. 11, pp. 231–247).

Guasti, M., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes, 20*(5), 667–696.

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.

Heyman, T., & Schaeken, W. (2015). Some differences in some: Examining variability in the interpretation of scalars using latent class analysis. *Psychologica Belgica, 55*(1), 1–18.

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review, 18*(4), 625–658.

Hirschberg, J. L. B. (1985). *A theory of scalar implicature.* Philadelphia: University of Pennsylvania.

Horn, L. R. (1972). *On the semantic properties of the logical operators in English.* UCLA: Doctoral diss.

Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.). *Georgetown University round table on languages and linguistics 1984* (pp. 11–42). Washington, D.C.: Georgetown University Press.

Hunt, L., Politzer-Ahles, S., Gibson, L., Minai, U., & Fiorentino, R. (2013). Pragmatic inferences modulate N400 during sentence comprehension: Evidence from picture–sentence verification. *Neuroscience Letters, 534*, 246–251.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122–149.

Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition, 124*(1), 66–71.

Kang, O., Rubin, D. O. N., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal, 94*(4), 554–566.

Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition, 120*(1), 67–81.

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences, 104*(30), 12577–12580.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*(2), 149.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647.

Lev-Ari, S. (2015). Comprehending non-native speakers: Theory and evidence for adjustment in manner of processing. *Frontiers in Psychology, 5*(1546), 1–12.

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.

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.

Levinson, S. C. (2000). *Presumptive meanings: the theory of generalized conversational implicature.* Cambridge, MA: MIT Press.

Mazzarella, D. (2015). Politeness, relevance and scalar inferences. *Journal of Pragmatics, 79*, 93–106.

Milroy, L., & Muysken, P. (1995). *One speaker, two languages: cross-disciplinary perspectives on code-switching.* Cambridge, UK: Cambridge University Press.

Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching languages, switching palabras (words): An electrophysiological study of code switching. *Brain and Language, 80*(2), 188–207.

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*(3),

289–306.
Myers-Scotton, C., & Jake, J. L. (2000). Testing the 4-M Model: Introduction. *International Journal of Bilingualism, 4*(1), 1–8.
Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science, 13*(4), 329–336.
Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience, 22*(12), 2886–2898.
Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18*(1), 62–85.
Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language, 63*(3), 324–346.
Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition, 78*(2), 165–188.
Noveck, I. A., & Posada, A. (2003). Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language, 85*(2), 203–210.
Ozturk, O., & Papafragou, A. (2015). The acquisition of epistemic modality: From semantic meaning to pragmatic interpretation. *Language Learning and Development, 11*(3), 191–214.
Ozturk, O., & Papafragou, A. (2016). The acquisition of evidentiality and source monitoring. *Language Learning and Development, 12*(2), 199–230.
Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology, 66*(2), 232–258.
Pavlenko, A. (Ed.). (2009). *The bilingual mental lexicon: Interdisciplinary approaches*. Bristol, UK: Multilingual Matters.
Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(04), 329–347.
Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29*(2), 191–215.
Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en Espanol: Toward a typology of code-switching. *Linguistics, 18*(2), 221–256.
Prat, C. S., & Just, M. A. (2011). Exploring the neural dynamics underpinning individual differences in sentence comprehension. *Cerebral Cortex, 21*(8), 1747–1760.
Prat, C. S., Keller, T. A., & Just, M. A. (2007). Individual differences in sentence comprehension: A functional magnetic resonance imaging investigation of syntactic and lexical processing demands. *Journal of Cognitive Neuroscience, 19*(12), 1950–1963.
Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition, 13*(02), 253–262.
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.
Romero-Rivas, C., Martin, C. D., & Costa, A. (2015). Processing changes when listening to foreign-accented speech. *Frontiers in Human Neuroscience, 9*(167), 1–15.
Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W. M., & Werker, J. F. (2012). A bilingual advantage in visual language discrimination in infancy. *Psychological Science, 23*(9), 994–999.
Sedivy, J. C. (2007). Implicature during real time conversation: A view from language processing research. *Philosophy Compass, 2*(3), 475–496.
Skordos, D., & Papafragou, A. (2016). Children's derivation of scalar implicatures: Alternatives and relevance. *Cognition, 153*, 6–18.
Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language, 25*(4), 359–393.
Sperber, D., & Wilson, D. (1986). *Relevance: communication and cognition.* Blackwell.
Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia, 56*, 289–301.
Tavano, E., & Kaiser, E. (2010). Processing scalar implicature: What can individual differences tell us? *University of Pennsylvania Working Papers Linguistics, 16*(1), 24.
Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*(4), 580–591.
Van Heugten, M., & Johnson, E. K. (2014). Learning to contend with accents in infancy: Benefits of brief speaker exposure. *Journal of Experimental Psychology: General, 143*(1), 340–350.
Wei, L. (2002). The bilingual mental lexicon and speech production process. *Brain and Language, 81*(1), 691–707..
Zhao, M., Liu, T., Chen, G., & Chen, F. (2015). Are scalar implicatures automatically processed and different for each individual? A mismatch negativity (MMN) study. *Brain Research, 1599*, 137–149.