Original Article

QJEP

Quarterly Journal of Experimental Psychology
2019, Vol. 72(4) 764–778

DOI: 10.1177/1747021818765547
qjep.sagepub.com
SAGE

Speaker-specific processing of anomalous utterances

Laurel Brehm[1,2,3], Carrie N Jackson[1,4] and Karen L Miller[1,5]

Abstract

Existing work shows that readers often interpret grammatical errors (e.g., *The key to the cabinets *were shiny)* and sentence-level blends ("without-blend": *Claudia left without her headphones *off*) in a non-literal fashion, inferring that a more frequent or more canonical utterance was intended instead. This work examines how interlocutor identity affects the processing and interpretation of anomalous sentences. We presented anomalies in the context of "emails" attributed to various writers in a self-paced reading paradigm and used comprehension questions to probe how sentence interpretation changed based upon properties of the item and properties of the "speaker." Experiment 1 compared standardised American English speakers to L2 English speakers; Experiment 2 compared the same standardised English speakers to speakers of a non-Standardised American English dialect. Agreement errors and without-blends both led to more non-literal responses than comparable canonical items. For agreement errors, more non-literal interpretations also occurred when sentences were attributed to speakers of Standardised American English than either non-Standardised group. These data suggest that understanding sentences relies on expectations and heuristics about which utterances are likely. These are based upon experience with language, with speaker-specific differences, and upon more general cognitive biases.

Keywords

Sentence processing; subject–verb agreement; implicit negation; sentence blends; noisy channel; good-enough processing

Received: 15 May 2017; revised: 8 February 2018; accepted: 22 February 2018

Introduction

When an individual makes a speech error in conversation, their interlocutor can "listen through" the utterance, extracting an inferred meaning without difficulty even if the intended meaning and the utterance's literal meaning conflict (e.g., *I really like to—hate to get up in the morning*, Fromkin, 1971; *They iced the road* for *They salted the road*, Heidi Lorimor, personal communication, 2016). Similarly, when an individual uses a word or grammatical construction that is novel to their interlocutor *(That hoagie needs toasted*), the interlocutor often succeeds at inferring the intended meaning *(That sandwich needs to be toasted*). The implication is that listeners can interpret overt errors and novel variations by drawing upon past experience and expectations about what utterances are likely to be said. This work examines the processing of anomalous sentences, examining the end interpretation of anomalies that represent speech errors and variable forms (defined as probabilistic variants of two forms, for example, *ain't* vs *am not; needs done* vs *needs to be done*).

We focus on the question of speaker-specificity, examining how individuals interpret language anomalies attributed to native (L1) Standardised American English speakers, second-language (L2) American English speakers, and speakers from a regional, non-Standardised dialect of American English. The motivation for this study is

[1]Center for Language Science, The Pennsylvania State University, University Park, PA, USA
[2]Department of Psychology, The Pennsylvania State University, University Park, PA, USA
[3]Max Planck Institute for Psycholinguistics, Max-Planck-Gesellschaft, Nijmegen, The Netherlands
[4]Department of Germanic & Slavic Languages and Literatures, The Pennsylvania State University, University Park, PA, USA
[5]Department of Spanish, Italian, and Portuguese, The Pennsylvania State University, University Park, PA, USA

Corresponding author:
Laurel Brehm, Max Planck Institute for Psycholinguistics, Max-Planck-Gesellschaft, P.O. Box 310, 6500 AH Nijmegen, The Netherlands.
Email: laurel.brehm@gmail.com

that speakers from these three groups might vary in what type of variable forms and errors they produce. L2 speakers use variable forms unpredictably (see, for example, Hudson Kam & Newport, 2009) and produce errors unpredictably (see, for example, White, 2003), while the variable forms in different dialects of English might vary predictably (see, for example, Labov, 1963, 1972) and the types of errors produced in different dialects might be similar (as evidenced by similar generalisations from British and American data, see, for example, Cutler, 1982).

The basis for the current experiments is an emerging body of work suggesting that the message comprehenders draw from an utterance is not necessarily the literal meaning conveyed by the utterance's form. The finding is that comprehenders use their previous experience to infer meaning from anomalous sentences, seeming to "correct" implausible utterances (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Gibson, Bergen, & Piantadosi, 2013; Levy, 2008) and adapting quickly to morphosyntactic variability (e.g., Fine, Jaeger, Farmer, & Qian, 2013; Fraundorf & Jaeger, 2016). The premise of this study is to inform how experience and expectations affect anomaly comprehension, asking whether errors and variable forms are interpreted differently depending on properties of the ostensible speaker. We begin with an overview of recent work on non-literal language comprehension and then turn to the question of how individuals predict and adapt to L1 and L2 variability in language.

Noisy-channel and good-enough processing

To date, work describing non-literal comprehension of sentences has fallen under two broad frameworks—*noisy channel* comprehension and *good-enough processing*. Work within these frameworks provides converging evidence on the dynamics of comprehension and the maintenance of uncertainty during processing. Both frameworks capture the fact that readers often interpret an anomalous utterance non-literally by drawing upon experience and heuristics about what utterances are likely.

The noisy channel framework seeks to describe why individuals do not always veridically interpret language. The phenomenon is that when faced with a semantically implausible sentence (e.g., *The mother gave the daughter to the candle;* Gibson et al., 2013) or a common grammatical error (e.g., *The key to the cabinets *were shiny*; Patson & Husband, 2016), readers often "correct" the sentence, interpreting it as if it were semantically plausible (*The mother gave the daughter the candle*) or grammatically correct (*The keys to the cabinets were shiny*). Responses like these reflect non-literal interpretations of the presented utterance.

Importantly, rates of non-literal interpretations increase in the presence of a source of variability ("noise") to which an error in the reader's comprehension can be attributed (e.g., Gibson et al., 2013), such that implausible utterances spoken by L2 individuals elicit more corrected non-literal interpretations than those produced by L1 individuals (Gibson et al., 2017).

Patterns of utterance reinterpretation also follow expectations based on prior language experience and the content of the sentence as it unfolds (e.g., Levy, 2008; Levy, Bicknell, Slattery, & Rayner, 2009). The implication is that comprehenders may model the likely errors of their interlocutors, recovering what they infer to be the speaker's intended meaning in the face of plausible speech errors (cast in terms of edit distance, Gibson et al., 2013; inferred sentence blends/doubling of words, Frazier & Clifton, 2011, 2015; inferred subject–verb agreement errors, Patson & Husband, 2016). This underscores the importance of modelling a speaker's *intent* in addition to tracking what the speaker has actually said.

In addition to recasting the form of anomalous utterances based upon more likely alternatives, readers also sometimes construct a *good enough* or heuristically driven interpretation of a sentence. For instance, *While Anna dressed the baby played in the crib* can imply the non-syntactically licenced meaning "Anna dressed the baby" (Christianson et al., 2001). Linguistic and real-world experience influences these misinterpretations, such that plausibility and word order heuristics drive readers' interpretations of what they have read (Christianson, Luke, & Ferreira, 2010; Ferreira, 2003). This underscores how expectations about intended meanings and intended forms support comprehension.

Prediction and adaptation to L1 and L2 variability

Processing anomalous utterances might be recast as processing in the face of variability, such that an individual is faced with extracting meaning from an utterance that differs from what is expected. We draw upon work on tracking speaker-specific variability to provide further motivation for predictions about non-literal processing of L1, L2, and dialectal utterances.

A large body of work demonstrates that comprehenders track variability across many levels of language. Experimental studies show that readers adapt quickly to written non-Standardised forms (e.g., Fraundorf & Jaeger, 2016; Kaschak & Glenberg, 2004), and that both anomalous utterances (Ivanova, Branigan, McLean, Costa, & Pickering, 2017) and novel non-Standardised forms (Kaschak, 2006) facilitate later processing of similar utterances. These findings show that with as few as 10 trials of exposure to a systematic anomaly or variable form (as used in the training phase in Kaschak & Glenberg, 2004), individuals adapt to the novel form.

We posit that systematicity is a key reason for individuals' swift adaptation to the anomalies or non-Standardised forms produced by L1 speakers. Native speakers make relatively few speech errors and the errors they make follow predictable patterns (see, for example, Fromkin, 1971 for an overview). The variable forms produced by native speakers in their regional dialects are often socially conditioned and contextually dependent (e.g., Labov, 1963, 1972). This means that the L1 production of non-Standardised utterances is relatively predictable.

In contrast, while adaptation to variability in L2 speech does occur, it is often effortful and relatively slow. Work using time-sensitive techniques (eye-tracking, event-related potential [ERP]) shows that L1 individuals listening to L2 speech make fewer predictions about the morphosyntactic (e.g., Hanulíková, Van Alphen, Van Goch, & Weber, 2012) properties of non-native speech and draw fewer inferences about non-native speaker disfluencies (Bosker, Quené, Sanders, & de Jong, 2014) and gender agreement errors (Hopp, 2016).

Adaptation to variability in L1 and L2 speech alike seems to be dependent on experience, such that experience with specific speakers and specific accents facilitates comprehension (see, for example, Bradlow & Bent, 2008; Nygaard & Pisoni, 1998; Kennedy & Trofimovich, 2008). This suggests that the weaker, less consistent predictions made for L2 utterances could be due to the lack of a specific inference to be made about the intended meaning or form. Following this logic, Grey and van Hell (2017) showed that predictions about L2 speech are influenced by the perception of who is talking. Only individuals who correctly identified a foreign-accented speaker as L1 Chinese showed a strong neural response to a grammatical error typical of this group (pronoun gender error), suggesting that participants' prior exposure to speakers of this group influenced expectations during processing. The implication is that predictions are made based upon experience with indexical features of the speaker, meaning that processing of L2 anomalies may be speaker-specific and dependent upon the comprehender's experience with speakers of the relevant group. This provides a link between non-literal processing frameworks and adaptation to variability in L1 and L2 utterances.

Speaker-specific anomaly processing

The work reviewed above on non-literal comprehension shows that language is processed based upon probabilistic, heuristic-driven inferences about likely utterances, while the work reviewed on regional dialect and L2 processing suggests that individuals track information specific to groups of speakers and can deploy this information strategically in processing. This research sits at the intersection of these two domains.

The current studies examine the processing of two types of sentence-level anomalies attested in real-world production: *subject–verb agreement* and *without-blends*. These constructions were selected to reflect an example of an error and a variable, non-Standardised form, with the goal of assessing common mechanisms of speaker-specific processing across different types of anomalies. We review properties of these constructions below.

Subject–verb agreement anomalies. Errors in subject–verb number agreement have been extensively studied in production, using sentence completion paradigms (e.g., Bock & Miller, 1991; Brehm & Bock, 2017; Lorimor, Jackson, Spalek, & van Hell, 2016), and in comprehension, using reading studies and measures of sentence misinterpretations (e.g., Patson & Husband, 2016; Pearlmutter, Garnsey, & Bock, 1999; Wagers, Lau, & Phillips, 2009). The basic finding is that in production, singular heads with plural local (intervening) nouns elicit more number-marking errors on the verb (productions of 1a vs 1b, conventionally correct) than do comparable items with singular local nouns (productions of 1c vs 1d). In comprehension, ungrammatical plural verbs following local plural nouns (1a) incur less of a slow-down in reading times as compared to those following singular local nouns (1c; Wagers et al., 2009). Plural verbs and plural nouns alike increase participants' tendency to interpret the head as having been plural in offline judgments, with increased non-literal "yes" responses to questions such as Q1 for items containing a plural noun or ungrammatical plural verb (1b and 1c: both about 23%), and with the most non-literal responses for items containing both a plural noun and ungrammatical plural verb (1a: about 40%, Patson & Husband, 2016):

1a. The sign on the taxis actually *were lit up brightly
1b. The sign on the taxis actually was lit up brightly
1c. The sign on the taxi actually *were lit up brightly
1d. The sign on the taxi actually was lit up brightly
Q1. Was there more than one sign?

In addition to eliciting processing and production difficulty, subject–verb agreement is highly variable within and between dialects. This suggests it is a construction where speaker-specific inferences and reader experience may be important for comprehension. Some varieties of English (including Appalachian English, as spoken in Western and Central Pennsylvania) allow utterances such as "*There was* four *of these houses*" (e.g., Tagliamonte & Baayen, 2012; Tortora & den Dikken, 2010). Meaning-based agreement variations are also widely present in standardised L1 English (Brehm & Bock, 2013, 2017) and the L2 English spoken by proficient L2 speakers (Foote, 2010; Jackson, Mormer, & Brehm, in press; Nicol & Greth, 2003). The pattern is that plural verbs are frequently produced with grammatically singular but conceptually plural referents (*The label on the bottles were*). This suggests that to a comprehender, a plural verb produced with a grammatically singular referent could be perceived as meaning-driven and not erroneous. This means that variation in agreement production is often multiply determined, leaving it to the listener to infer the presence of an intended variation versus a speech error.

The current literature also suggests that production and comprehension may recruit overlapping but distinct mechanisms (e.g., Lorimor, Jackson, & Foote, 2015; Tanner, Nicol, & Brehm, 2014). This dissociation between agreement comprehension and production has consequences for L2 speakers (see, for example, Trenkic, Mirkovic, & Altmann, 2014). The implication is that a comprehender's own processing difficulty and their inferences about *production* difficulty are both relevant. Unlike many other constructions previously investigated under non-literal processing frameworks, there are as such two clear ways to repair a subject–verb agreement anomaly. This stems from the fact that subject–verb agreement is a dependency between two elements. One repair is to change the inflection on the head noun (adding an *-s*), while the other is to change the inflection on the verb (replacing *were* with *was).* To the extent that a comprehender attributes the anomaly to their own miscomprehension, they may be more likely to infer that the head number should be changed; to the extent that a comprehender attributes the anomaly to the speaker's misproduction, they may be more likely to infer that the head number was produced as intended. The dissociation between repairs to be made for comprehension- and production-centred inferences sets these constructions apart from those used in previous work (e.g., "*The mother gave the candle the daughter*"; e.g., Gibson et al., 2013; Gibson et al., 2017), where there is only one parsimonious repair to be made (add *to*). This allows us to examine the unique contributions of reader miscomprehension and speaker misproduction to non-literal inferences.

Without-blends. Blends of two possible utterance plans are another type of production error, deriving from the simultaneous consideration of two different utterance formulations (e.g., Butterworth, 1982; Cutting & Bock, 1997). "Without" or "implicit negation" blends, such as *I just like the way the president looks without his shirt off* are a type of blend attested in natural speech which elicit highly variable comprehension (Frazier & Clifton, 2015). About 65% of the time items like *Claudia walked to class without her headphones off* (2a) are interpreted as reflecting a literal, compositional meaning ("Claudia had her headphones on"), whereas about 35% of the time, such items are interpreted as reflecting a negative concord or "unblended" meaning ("Claudia had her headphones off"). This means that a sentence such as 2a can be interpreted identically to a sentence such as 2b, eliciting the same answer to a question such as Q2: "No, Claudia was not wearing headphones":

2a. Claudia walked to class without her headphones off yesterday afternoon
2b. Claudia walked to class without her headphones on yesterday afternoon
Q2. Was Claudia wearing her headphones when she walked to class?

The variable interpretation of these without-blends may derive from a variety of sources, which situates this construction as another for which a comprehender's inference about the anomaly's origin could be particularly salient. As indicated by the terminology above, we suggest that these items might be treated by readers as a variable form similar to negative concord. Difficulty in producing and interpreting utterances like 2a may reflect the inherent complexity of negation (e.g., Horn, 1989). Negation, especially implicit negation (e.g., Evans, Clibbens, & Rood, 1996), involves several syntactic and semantic operations; this may increase error rates in the production of negation-containing utterances. This then requires the reader to infer whether the utterance contained an error or was produced as intended, leading to the two salient meanings of this utterance (see, for example, Frazier & Clifton, 2015).

The way that negation is expressed is also variable across dialects and languages, suggesting that processing of without-blends may be speaker-specific. For example, Spanish and English express negation differently, such that Spanish requires negative concord (using two negative elements to express a single negative meaning) and English has negative concord available only in certain casual speech registers and non-Standardised dialects, such as Appalachian English (see, for example, Blanchette, 2013; Bosque & Demonte, 1999). As with agreement anomalies, this suggests that the interpretation of a without-blend might depend on the inference as to *why* it was produced. If the reader assumes that the utterance 2a reflects a correct production, the response to a question such as Q2 would be a literal "yes" response. In contrast, if the reader assumes their own miscomprehension, the speaker's misproduction, or that the speaker is using a variable form as intended, the response would be a non-literal "no."

Without-blends differ in two other important ways from the agreement anomalies, informing some useful comparisons between the two sentence types. The first is in the type of non-literal inference that the reader is required to make. To obtain a non-literal interpretation of a without-blend, the reader only needs to change the construed meaning without making any changes to the morphosyntactic structure of the utterance. This may mean that non-literal processing of without-blends may not require a syntactic reanalysis. Contrasting the two sentence types therefore provides insight into whether semantic and syntactic aspects of language processing are differentially impacted by speaker-specific properties.

A second important difference between without-blends and agreement anomalies relies upon experience. Readers are less likely to have specific experience with individuals producing without-blends: while attested in the literature (and observed by the second author in conversation), without-blends are infrequent. In contrast, agreement errors are relatively common in running speech. As such, observing speaker-specific processing in without-blends versus agreement anomalies provides a strong test for the role of heuristics versus experience-driven predictions in non-literal processing.

Speaker-specific predictions and sources of variability. The current experiments use the non-literal interpretation of subject–verb agreement and without-blends to test what information is deployed in language comprehension. As reviewed above, these utterances were selected to reflect two different attested variants (a grammatical error requiring morphosyntactic reanalysis, a plausible variable form requiring semantic reanalysis) that have been shown to be variable in L2 speech and regional dialects of American English.

We draw two competing sets of predictions for speaker-specific anomaly interpretation. Existing literature shows that non-literal inferences occur, and that they are based upon expectations and biases about properties of the language system and the speaker's communicative intent (e.g., Ferreira, 2003; Gibson et al., 2013 among many). Contrasting different speaker groups allows us to show the mechanisms behind non-literal comprehension.

Experiment 1 tests whether individuals process anomalies differently depending on whether they appear in sentences attributed to an L1 Standardised American English or L2 speaker. If what is most critical for processing is the perceived level of variability or "noise" in the system, especially on the producer's end, the prediction is that sentences attributed to L2 individuals will elicit more non-literal sentence interpretations, as L2 speakers are likely to be highly variable in their production and to produce more speech errors. The correction of these errors would appear as non-literal sentence interpretations. The prediction is then for differences to be especially apparent for items that are plausible errors (the ungrammatical subject–verb agreement items; the without-off items). Support for this hypothesis comes from the increased rates of non-literally interpreted implausible utterances (e.g., Gibson et al., 2017) and lexical items (e.g., Lev-Ari & Keysar, 2012) that are spoken by L2 individuals, consistent with the attribution of anomalies to a speaker's misproduction.

In contrast, if what is most critical for processing is the ability to easily infer the speaker's intended meaning, the opposite pattern is predicted, such that sentences attributed to Standardised speakers will elicit more non-literal interpretations. This is because the language anomalies produced by Standardised speakers are likely to be more systematic and more predictable than those produced by L2 speakers, and therefore easier to infer the intended meaning from. This means that an implausible L1 utterance is likely to be interpreted non-literally, as expectations and heuristics can be deployed to recover what is perceived to be the intended meaning. In contrast, an L2 speaker might frequently misspeak, but if the locus of the error cannot be determined (on the verb vs the noun; on "without" vs "off"), then no repair should be made. Under this account, non-literal inferences are predicted to be more likely for Standardised speakers, with the largest difference for items containing conflicting information where multiple types of repairs could be made (especially the local plural subject–verb agreement items, but also the without-off items). Support for this hypothesis comes from evidence for the easy adaptation to L1 variability (e.g., Fraundorf & Jaeger, 2016; Kaschak, 2006; Kaschak & Glenberg, 2004) and more difficult adaptation to L2 variability (Hanulíková et al., 2012; Hopp, 2016). The implication is that anomalies may be easier to repair in the case that they were produced by a Standardised speaker, as compared to an L2 speaker.

Experiment 2 tests the differences between processing Standardised utterances and those of regional dialects, linking the way linguistic variability is perceived and processed for proficient L2 speakers as compared to non-Standardised dialects and further highlighting what types of variability matter in language comprehension. This experiment serves as a replication of the patterns observed in Experiment 1. It also allows us to test whether speakers of a non-Standardised dialect are treated similarly to highly proficient L2 speakers. Specifically, to the extent that participants treat the types of variability inherent to both speaker groups similarly and have equivalent experience with L2 speakers and speakers of non-Standardised dialects, the two groups may elicit comparable rates of non-literal interpretations.

Experiment 1

Experiment 1 compared the processing of two types of anomalous sentences (agreement violations, without-blends) attributed to either Standardised American English speakers or to L1 Spanish–L2 English speakers. The critical test was whether sentence interpretations were affected by the attributed speaker.

Methods

Participants. In total, 141 participants with IP addresses from the United States were recruited via Amazon Mechanical Turk to capture a participant sample that covered a range of regional American dialects and varied on educational background and language experience. Of these 141 individuals, 10 were excluded for having reported learning a second language before the age of 5 years, 1 was excluded for self-reporting a learning disability, and 7 were excluded for comprehension performance below 80% correct on unambiguous filler trials. In addition, the data from one participant were lost due to a server error and the data from the second run of one participant who completed the experiment twice were excluded. The remaining 120 individuals all reported English as their

Table 1. Participant demographics by experiment.

	Experiment 1, $n = 120$		Experiment 2, $n = 120$	
Gender	Female	57	Female	45
	Male	63	Male	75
Age (years)	$M = 30.17$, range 19-55		$M = 30.29$, range 18-52	
Years of formal education	$M = 15.02$, range 12-22		$M = 14.33$, range 2-20	
Number of locations lived in for 2 years or more	$M = 2.51$, range 1-7		$M = 2.14$, range 1-8	
Number of distinct American dialect regions lived in	$M = 1.49$, range 1-5		$M = 1.42$, range 1-5	
Number of languages spoken	$M = 1.17$, range 1-3		$M = 1.20$, range 1-3	
Level of Education	High school or less	54	High school or less	53
	Professional training	2	Professional training	5
	At least some college	50	At least some college	53
	At least some graduate	7	At least some graduate	3
	MA	7	MA	5
	PhD	0	PhD	1
Daily exposure to foreign languages	No	108	No	112
	Yes	11	Yes	7
	NA	1	NA	1
Daily exposure to foreign accents	No	100	No	107
	Yes	20	Yes	13
Weekly exposure to foreign accents	0-1	96	0-1	98
	2-5	19	2-5	12
	6-10	2	6-10	4
	Over 10	3	Over 10	3
Daily exposure to regional dialects	No	95	No	110
	Yes	24	Yes	10
	NA	1		
Weekly exposure to regional dialects	0-1	95	0-1	105
	2-5	16	2-5	11
	6-10	3	6-10	0
	Over 10	4	Over 10	4
	NA	2		

NA: not applicable.

native language, had no self-reported learning difficulties or uncorrected visual impairments, scored above 80% accuracy on comprehension questions for unambiguous filler trials and were aged 18 years or above. See Table 1 for participant demographics.

Equipment. The experiment was run on participants' computers with an in-browser script presented using IbexFarm (Drummond, 2013). Participants were instructed to use the F and J keys on their keyboard to make "yes" and "no" responses, respectively. Key-press latencies were collected and sent to the server upon completion of the experiment. At recruitment, participants were told that the purpose of the experiment was to read emails and answer questions about them.

Materials. Participants read 168 sentences, framed as portions of email messages from four different authors. Of these, 72 were critical sentences. Critical sentences followed two patterns. *Agreement* sentences (48 items) followed the pattern outlined in 1a-1d.[1] *Without-blend* sentences (24 items) followed the pattern outlined in 2a-2b. All sentences began with an additional two to four word phrase (e.g., *no really; He saw that; I can't believe that*). These provided a context for the critical phrase to make it a more realistic email message and also served to make the location of the sentences' critical regions less predictable. See Appendix A in the Supplementary Material for a complete list of critical items.

Filler sentences contained 12 attachment ambiguity items (modified from Swets, Desmet, Clifton, & Ferreira, 2008), 24 semantically ambiguous items (based on "prevent near culmination" items from Frazier & Clifton, 2015), as well as object- and subject-extracted relative clauses (six of each). Like the critical items, these were also preceded by two to four word phrases to provide context. The remaining fillers (48 items) were casually written sentences describing classes, friends, family, and social activities of college students; a subset of these sentences (16 items) contained typos, text abbreviations, un-capitalised proper nouns, or punctuation errors.

All items were preceded by an email introduction containing an image of the message's author, her name, and a subject-line. Subject lines in critical sentences provided some context about the affective content or the topic of the upcoming sentence (e.g., *RE: wow!*).

All items were followed by a comprehension question. The critical comprehension questions probed the interpretation of the sentence. Agreement questions asked whether "more than one" of the head noun was present (as in Q1 above). Without-blend questions asked whether the agent was wearing the optional clothing/accessory (as in Q2 above). A subset of the filler comprehension questions (24 items) queried the participant's interpretation of an ambiguous sentence, while the remainder had an unambiguous answer. Of these unambiguous fillers, six queried the agent number, and the rest queried events described in the sentence or the emotional state of the author. Of all of the questions combined, 54% of items had a likely "no" answer.

Blocks of trials were introduced by a speech sample from one of four different speakers. Two speakers were speakers of Standardised American English: "Madison Parker" was recorded by a woman from Ohio, and "Ashley Clark" was recorded by a woman from Maryland; neither speaker had a discernable regional accent as judged by the authors. The other two speakers represented second-language speakers of English ("María Pérez" and "Sofía Menéndez"); both were recorded by women originally from Mexico but currently living in Central Pennsylvania. Each sound clip was about 12 s long. Introductions were paired with head-shots (see Appendix B in the Supplementary Material).

Procedure. The experiment itself was divided into eight experimental blocks, each of which contained 21 sentences attributed to a single individual. Each block began by introducing the author, as described above. Each trial began with a screen containing a small image of the message's author, the author's first name, and a subject line displayed for 3,000 ms. Then, the message text appeared in a single line. Dashes indicated the number and length of words in the sentence; upon pressing the spacebar, one word at a time was revealed and the preceding word disappeared. A yes/no comprehension question was then presented. Trials were followed by a 500 ms inter-stimulus interval; after four blocks, the participant was given the opportunity to take a break. Following the completion of experiment, participants completed a demographics and language background questionnaire (see Appendix C in the Supplementary Material for survey; see Table 1 for demographic results).

Design and data analysis. Items were assigned to different lists such that each participant saw one version of each item with an equal number of items in each condition for each sentence type and each speaker. Items were arranged into a fixed order per block using the programme Mix (van Casteren & Davis, 2006). To control for order of presentation, experimental blocks were ordered in three counter-balanced arrangements following the constraints that all four speakers were presented in the first four blocks, no more than two blocks of the same type of speaker appeared consecutively, and there were at minimum two blocks intervening before the second repetition of a speaker. From these factors, (four item versions, two types of speakers, and three block orders), 24 different experimental lists were created; five participants viewed each list.

We originally collected a sample of 48 individuals; the result of this sample suggested that any effects of interest are of a small to medium size. A priori power calculations for a within-subjects repeated measures design using G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009) show that a sample size of 120 individuals would give us at least 80% power to observe an effect size $f = .255$ or larger.[2] To correct for increased type I error in performing sequential analyses, we follow the guidelines outlined in Lakens (2014) and use the Pocock boundary to set the critical p-value to .0294 such that the overall alpha level for both analyses remains .05. We report only the results of the analysis containing all 120 individuals.

Data analysis was performed in R (R Core Team, 2014) using the package lme4 (version, 1.1-11, Bates et al., 2014). Logistic regressions were performed separately for each sentence type using the optimizer "bobyqa." The dependent measure for all analyses was the odds of responding literally (*literal* = 1, *non-literal* = 0). For agreement items, a non-literal response was a "yes" for all conditions, as previous work (Patson & Husband, 2016) has shown "no" to be the normative response. For without items, a non-literal response was "yes" to "without-on" items such as 2b and "no" to "without-off" items such as 2a. This coding was again based upon the normative responses in previous work (Frazier & Clifton, 2015).[3]

All two-level experimental factors were effects coded with contrasts (0.5, −0.5); this included Speaker (L2 vs Standardised), the sentence-level predictors in Agreement sentences (Local Noun: Singular vs Plural; Verb: Singular vs Plural) and the sentence-level predictors in Without-blend sentences (Preposition: On vs Off). This means that positive beta coefficients indicate more non-literal responses (and fewer literal responses) for Standardised speakers, plural nouns and verbs, and the preposition "off." In light of recent research showing rapid adaptation in reading comprehension (e.g., Fine et al., 2013), Block was also entered as a continuous predictor, centred such that zero represented performance at the mid-point of the experiment. Trials for which the comprehension question was answered extremely slowly (above 12 s) were excluded (0.04% of the data).

Random effects were fitted to the maximum structure justified by the experimental design. The maximum effect

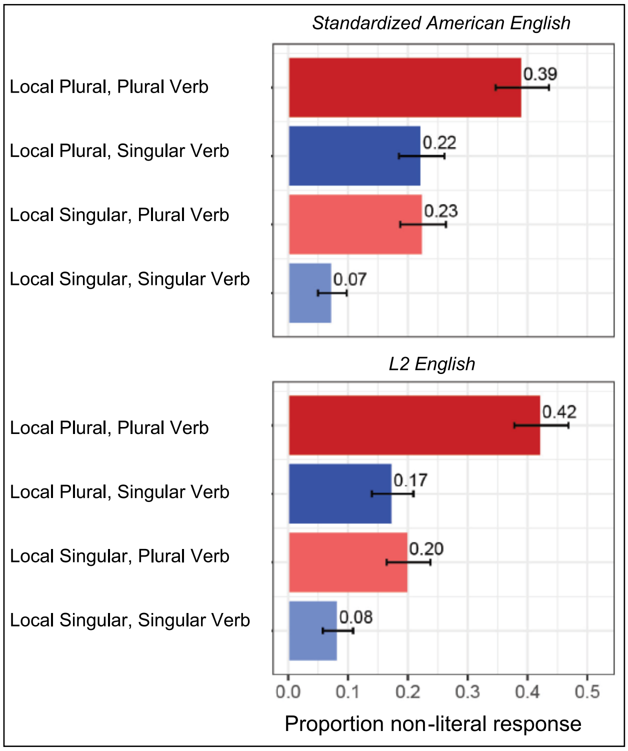

Figure 1. Agreement sentence interpretations for Experiment 1. Panels reflect sentences attributed to a Standardised American English speaker (top) and an L2 English speaker (bottom). Means by condition displayed next to bars; error bars reflect standard error around means.

Table 2. Results from logistic mixed-effect regression (log odds literal/non-literal response to critical questions) for Experiment 1 agreement trials.

Fixed effects	Estimate	*SE*	$p(\chi^2)$
Intercept	1.87	0.15	<.001
Block	0.07	0.03	.01
Local Noun	1.30	0.17	<.001
Verb	1.42	0.12	<.001
Speaker	0.10	0.09	.31
Local Noun × Verb	0.04	0.18	.81
Local Noun × Speaker	−0.08	0.17	.61
Verb × Speaker	0.14	0.17	.43
Local Noun × Verb × Speaker	−0.97	0.32	<.001
Random effects	**Groups**	***SD***	
Subject	Intercept	1.10	
	Local Noun	0.83	
	Speaker	0.36	
	Verb	0.87	
Item	Intercept	0.62	
	Local Noun	0.74	

SE: standard error; *SD*: standard deviation.

structure included random intercepts for subjects and items and random slopes by subjects and items for all main effects and interactions between Speaker and the sentence-level predictors for both sentence types. Random slopes were removed due to non-convergence to fit the maximum model justified by the data, and random slopes correlated above 0.95 were removed to avoid over-fitting. All *p*-values were calculated by model comparison (using chi-square tests).

Results

We report only the results of responses to comprehension questions in the main text. Reading time results can be found in Appendix D in the Supplementary Material.

Agreement items. See Figure 1 for a graphical depiction of non-literal ("yes, more than one") response rates; see Table 2 for results of mixed-effect models. A main effect of Local Noun was observed such that singular-headed sentences containing local plural nouns ("*taxis*," 1a and 1b) elicited more non-literal plural interpretations (Local plural $M = 30\%$ vs Local singular $M = 15\%$). Similarly, a main effect of Verb was observed such that singular-headed sentences containing ungrammatical plural verbs ("*were*," 1a and 1c) also elicited more non-literal plural interpretations (ungrammatical $M = 31\%$ vs grammatical $M = 14\%$). These main effects were qualified by an interaction between Local Noun, Verb, and Speaker (see Figure 1 for condition means). The pattern was such that items with mismatching local nouns and verbs (1b: "*the sign on the taxis was*"; 1c: "*the sign on the taxi were*") led to more non-literal responses for the Standardised speaker than the L2 speaker, while items with plural local nouns and plural verbs (1a: "*the sign on the taxis were"*) led to fewer non-literal responses for the Standardised speaker than the L2 speaker. A main effect of Block was also observed such the rate of non-literal responses decreased throughout the experiment (from 29% in block 1 to 19% in block 8). No other effects approached significance.[4]

Without-blends. See Figure 2 for a graphical depiction of literal vs non-literal response rates and see Table 3 for mixed-effect models. Items containing the potentially anomalous "*off*" elicited significantly more non-literal responses (41%) than did items containing "on" did (7%), as evidenced by a significant main effect of Preposition type. There was also an effect of Block such that the rate of non-literal responses decreased throughout the experiment (from 34% in block 1 to 20% in block 8). No effects including Speaker were significant and no other effects approached significance.

Discussion

Experiment 1 was designed to examine the interpretation of two types of anomalous sentences. The first type of

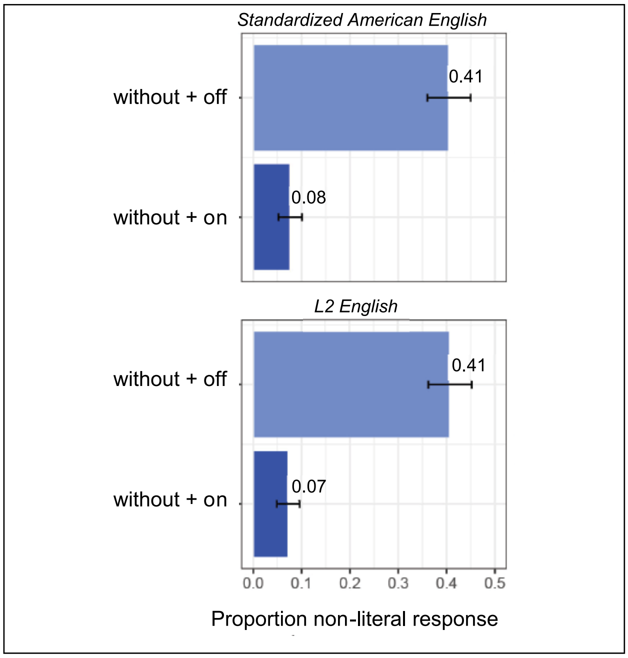

Figure 2. Without-blend sentence interpretations for Experiment 1. Panels reflect sentences attributed to a Standardised American English speaker (top) and an L2 English speaker (bottom). Means by condition displayed next to bars; error bars reflect standard error around means.

sentence contained a subject–verb agreement anomaly. Ungrammatical sentences that were likely to reflect a production error (e.g., *The sign on the taxis actually *were . . .*) elicited many non-literal responses to questions such as "*Was there more than* one *sign?*," while grammatical sentences unlikely to reflect any error or to induce any processing difficulty (*The sign on the taxi actually was . . .*) elicited few non-literal responses. This replicates previous work (Patson & Husband, 2016). In addition, the likelihood of non-literal responses also dropped throughout the course of the experiment, showing a role for adaptation to within-experiment language statistics.

Agreement items containing mismatching local nouns and verbs elicited differing rates of non-literal responses depending on the speaker they were attributed to. Items containing a local plural noun and a singular verb (*The sign on the taxis actually was . . .*) and items containing a local singular noun and a plural verb (*The sign on the taxi actually *were . . .*) both led to more non-literal responses for a Standardised speaker than an L2 speaker. This shows the importance of transparency of the speaker's intended meaning above the likelihood of speaker error in eliciting non-literal responses to subject–verb agreement. The specific pattern reflects the greater use of noun and verb number cues for Standardised versus L2 individuals, even to the extent that noun number cues can lead to overwriting a verb that appropriately matches the head. It is consistent with the fact that L2 individuals are less likely to be

Table 3. Results from logistic mixed-effect regression log odds literal/non-literal response to critical questions for Experiment 1 without-blend trials.

Fixed effects	Estimate	*SE*	$p(\chi^2)$
Intercept	2.05	0.19	<.001
Block	0.22	0.04	<.001
Preposition	2.89	0.28	<.001
Speaker	0.04	0.13	.79
Preposition × Speaker	0.09	0.27	.72
Random effects	Groups	*SD*	
Subject	Intercept	1.49	
	Preposition	1.85	
Item	Intercept	0.40	

SE: standard error; *SD*: standard deviation.

reliable in their use of number inflections, suggesting that the non-literal inference could be based on sensitivity to observed differences between Standardised and L2 speech.

The second type of sentence involved a plausible blend of implicitly negative elements. Without-blend sentences (e.g., *without her headphones off*) elicited more non-literal interpretations than unblended control sentences (e.g., *without her headphones on*), consistent with previous work showing that the potentially anomalous preposition *off* is often interpreted as a blend or negative concord (Frazier & Clifton, 2015). Again, rates of non-literal responses dropped throughout the course of the experiment, suggesting a role for adaptation within the experiment. However, no differences were observed for Standardised speakers as compared to L2 speakers. The lack of speaker-specific inferences might follow from the attribution of this anomaly to semantic heuristics and/or general planning difficulty, in contrast to experience with language produced by specific groups of speakers.

Experiment 2

Experiment 1 showed that sentences containing potential anomalies were likely to be interpreted non-literally. This occurred for subject–verb agreement and without-blends. Experiment 1 also showed that subject–verb agreement anomalies elicited more non-literal interpretations when attributed to a Standardised speaker than an L2 speaker. This points to a speaker-specific effect such that more non-literal interpretations occur when the speaker's message is perceived as more transparent by the reader. To replicate and extend these results, Experiment 2 used a non-Standardised speaker that represented a regional dialect of American English (rural Pennsylvania English). This is designed to query the nature of the relevant speaker-specific properties for comprehension. If the relevant property for predicting a speaker's intended meaning is that they use Standardised American English, a variant of

English that is likely to be familiar to all participants, the prediction is that a non-Standardised variety of English should show results similar to L2 English speakers, such that Standardised speakers lead to more non-literal interpretations in Experiment 2. In contrast, if the relevant property is perceived nativeness or an expectation of fluency, the prediction is that a non-Standardised variety of English should be similar to a Standardised variety, as both groups may be perceived as native, fluent speakers.

Methods

Participants. In total, 144 individuals with IP addresses from the United States were recruited with Amazon Mechanical Turk. Of these 144 individuals, 15 were excluded for reporting learning a second language before the age of 5 years, 4 were excluded due to self-reported learning difficulties, and 4 were excluded for comprehension accuracy below 80% on unambiguous filler trials. The second run of one participant who participated in the experiment twice was also excluded. The remaining 120 individuals all reported English as their native language, had no self-reported learning difficulties or uncorrected visual impairments, scored above 80% on comprehension questions for unambiguous filler trials, and were 18 years or older. See Table 1 for participant demographics.

Equipment. Identical to Experiment 1.

Materials. The standardised American English speakers were the same as in Experiment 1. The other two speakers represented individuals from Central Pennsylvania. These speakers ("Shelby Williams" and "Tiffany Johnson") were recorded by women from Central Pennsylvania (see Appendix B in the Supplementary Material). This dialect was selected to avoid using a stigmatised variety of American English (e.g., Southern American English) and because it represents the dialect spoken in the authors' local environment. To highlight potential differences to the Standardised speakers, the non-Standardised speaker introductions contained information about being from a small town (see Appendix A in the Supplementary Material); we therefore refer to this group as "Rural Pennsylvanian." Both speakers had a noticeable regional accent, as judged by the authors, and were portrayed with different head-shots than the L2 speakers from Experiment 1; all other materials were identical.

Procedure. Identical to Experiment 1.

Design and data analysis. Identical to Experiment 1. As in Experiment 1, we originally collected a sample of 48 individuals and increased the sample size to 120 individuals after observing that any effects of interest were of a small to medium size. To correct for increased type I error, we use the Pocock boundary to set the critical p-value to .0294

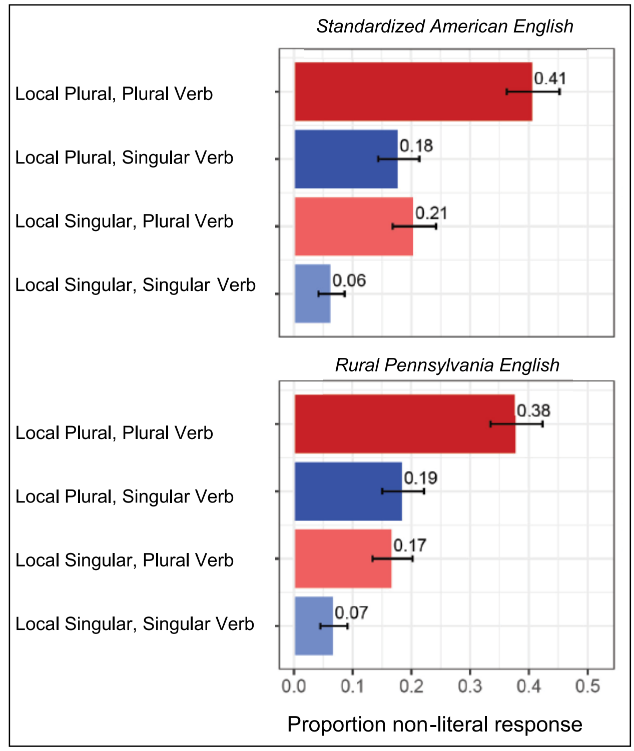

Figure 3. Agreement sentence interpretations for Experiment 2. Panels reflect sentences attributed to a Standardised American English speaker (top) and a non-Standardised dialect speaker (bottom). Means by condition displayed next to bars; error bars reflect standard error around means.

and report only the results of the second analysis containing all 120 individuals.

As in Experiment 1, trials for which the comprehension question was answered extremely slowly (above 12 s) were excluded (2% of the data).

Results

We report only the results of responses to comprehension questions in the main text. Reading time results can be found in Appendix D in the Supplementary Material.

Agreement items. See Figure 3 for a graphical depiction of no (literal)/yes (non-literal) response rates; see Table 4 for results of mixed-effect models. Main effects of Local Noun and Verb were observed such that sentences containing local plural nouns ("*taxis*," 1a and 1b) elicited more non-literal plural responses (Local plural M=29% vs Local singular M=13%), as did sentences containing ungrammatical plural verbs ("*were*," 1a and 1c; Ungrammatical M=29% vs Grammatical M=12%); there was again no reliable interaction between the two factors. These effects were qualified by an interaction between Verb and Speaker such that ungrammatical plural verbs led to more non-literal plural responses for the Standardised

Table 4. Results from logistic mixed-effect regression (log odds literal/non-literal response to critical questions) for Experiment 2 agreement trials.

Fixed effects	Estimate	*SE*	$p(\chi^2)$
Intercept	2.13	0.16	< 0.001
Block	0.09	0.03	< 0.01
Local Noun	1.47	0.19	< 0.001
Verb	1.39	0.16	< 0.001
Speaker	0.02	0.10	0.87
Local Noun × Verb	−0.15	0.19	0.46
Local Noun × Speaker	0.04	0.18	0.84
Verb × Speaker	−0.44	0.17	0.01
Local Noun × Verb × Speaker	−0.26	0.33	0.45
Random effects	**Groups**	***SD***	
Subject	Intercept	1.21	
	Local Noun	1.00	
	Verb	0.96	
	Speaker	0.27	
Item	Intercept	0.64	
	Local Noun	0.84	
	Verb	0.51	

SE: standard error; *SD*: standard deviation.

speaker than the Rural Pennsylvania speaker (for plural verbs, Standardised $M=31\%$ vs PA $M=27\%$, for singular verbs, Standardised $M=12\%$ vs PA $M=13\%$). An effect of Block was also observed such that the rate of non-literal responses decreased throughout the course of the experiment (from 28% in block 1 to 19% in block 8). No other effects approached significance.

Without-blends: responses. See Figure 4 for a graphical depiction of literal vs non-literal response rates and see Table 5 for mixed-effect models. We again observed a significant effect of Preposition such that participants were more likely to respond non-literally for sentences containing the potentially anomalous "*off*" (2a, $M=41\%$) as compared to sentences containing "*on*" (2b, $M=5\%$). There was also an effect of Block such that the rate of non-literal responses decreased throughout the course of the experiment (from 34% in block 1 to 17% in block 8). No other significant differences were observed.

Discussion

Results of Experiment 2 were consistent with those of Experiment 1. For the first sentence type, involving subject–verb agreement anomalies, items likely to reflect an error (*The sign on the taxis *were*) received frequent non-literal plural interpretations, while items unlikely to reflect an error (*The sign on the taxi was*) received infrequent non-literal interpretations; this replicates Experiment 1 and Patson and Husband (2016). There was also again an effect of Block such that non-literal responses decreased through the course of the experiment. This demonstrates a role for adaptation to the language environment of the experiment.

We also observed an interaction between Speaker and Verb such that sentences containing ungrammatical plural verbs elicited more non-literal plural responses for the Standardised speaker than the Rural Pennsylvania speaker.

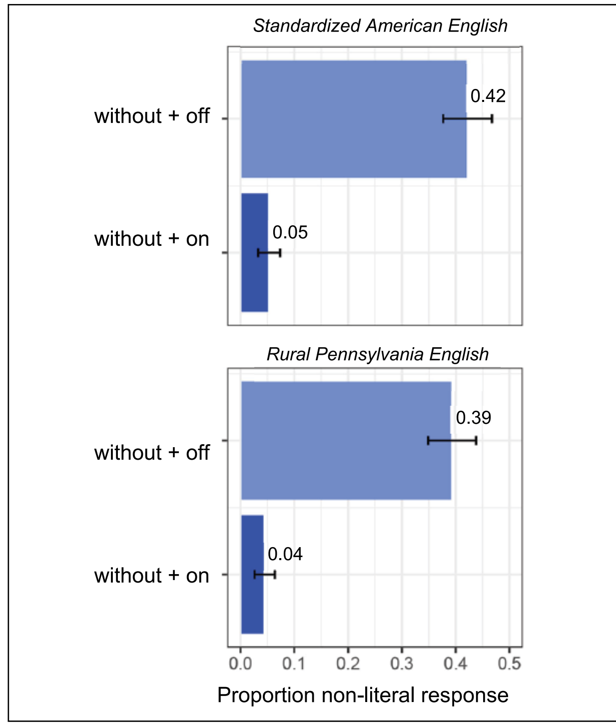

Figure 4. Without-blend sentence interpretations for Experiment 2. Panels reflect sentences attributed to a Standardised American English speaker (top) and a non-Standardised dialect speaker (bottom). Means by condition displayed next to bars; error bars reflect standard error around means.

Table 5. Results from logistic mixed-effect regression (log odds literal/non-literal response to critical questions) for Experiment 2 without-blend trials.

Fixed effects	Estimate	SE	$p(\chi 2)$
Intercept	2.12	0.18	<.001
Block	0.28	0.04	<.001
Preposition	3.14	0.29	<.001
Speaker	0.20	0.15	.18
Preposition × Speaker	−0.02	0.29	.96
Random effects	**Groups**	***SD***	
Subject	Intercept	1.27	
	Preposition	1.78	
Item	Intercept	0.37	

SE: standard error; *SD*: standard deviation.

The interaction between Speaker and Verb is in the same direction as Experiment 1, suggesting the importance of a transparently recovered message over and above the likelihood of an individual producing an error for non-literal responding. The implication is that assuming the speaker belongs to a Standardised American English group facilitates a non-literal inference that changes the sentence's meaning and form. Proficient L2 English and a regional dialect are treated similarly, and both lead to differences in how subject–verb agreement is understood.

The form of the interaction between Speaker and Verb is also consistent with the reduced predictability of the verb number from the head number in certain regional dialects of American English. In particular, we note that Appalachian American English is less reliable on its verb inflection than Standardised American English, such that plural heads can govern grammatically singular verbs (e.g., utterances like *We was going* are allowed, Tortora & den Dikken, 2010). This might cause verb cues to be less reliable for Rural Pennsylvania English, leading to less frequent non-literal inferences based upon verb number. This suggests a plausible role of experience with different speaker groups in deriving non-literal interpretations.

The second type of anomaly, without-blends, also replicated Experiment 1. Anomalous without-blend items like "*without her headphones off*" were often interpreted non-literally, with a main effect of Preposition such that non-literal responses were more likely when the item contained a potential anomaly. Non-literal responses also decreased throughout the course of the experiment, suggesting a role for within-experiment experience, but there was no effect of Speaker, suggesting a limited role for speaker-specific experience in comprehending anomalies like without-blends. As in Experiment 1, this suggests that non-literal processing of without-blends arises due to application of semantic heuristics and/or general planning or processing difficulty, factors that would not necessarily be expected to vary between speaker groups.

General discussion

In two experiments, we showed evidence of non-literal processing of plausibly anomalous utterances that reflected attested production variability. Individuals made more non-literal plural interpretations of sentences containing plural verbs and plural nouns (vs those with singular nouns and singular verbs) and more non-literal "unblended" interpretations of plausible implicit negation blends (vs control stimuli). This is consistent with the finding that comprehenders interpret utterances following what is inferred to be likely rather than what has literally been observed, as outlined by noisy channel and good-enough processing frameworks (e.g., Christianson et al., 2001, 2010; Ferreira, 2003; Gibson et al., 2013; Gibson et al., 2017; Frazier & Clifton, 2015; Levy, 2008). The agreement findings replicate previous work (Patson & Husband, 2016), while the without-blend findings demonstrate that responses to comprehension questions mirror semantic judgments (Frazier & Clifton, 2015). The implication is that heuristics, biases, and expectations drawn from previous experience (or lack of experience) impact interpretation.

We also found that for agreement anomalies, rates of non-literal interpretations were dependent on the ostensible identity of the sentence's author. Items attributed to Standardised speakers of American English containing mismatching local noun and verb number led to more non-literal plural interpretations than the same items attributed to L2 individuals. This suggests that mismatching number cues are perceived as less consistent with the intended message when attributed to Standardised speakers than to L2 speakers, who may be idiosyncratic in their use of number inflections. Similarly, items attributed to Standardised speakers that contained an anomalous plural verb led to more non-literal plural interpretations than the same items attributed to individuals from a regional dialect of American English, who may be idiosyncratic in their use of verb number. These patterns suggest that what matters for non-literal responding is the ease with which a comprehender believes they can infer the speaker's intended message, changing what is literally present into something that is perceived as more probable when the utterance contains a plausible error. These non-literal response patterns are consistent with changes in comprehension based upon speaker-specific expectations about morphosyntax, consistent with the noisy channel and good-enough processing frameworks. These patterns also underscore that sufficiently proficient L2 speakers are treated similar to speakers from a regional dialect of American English.

On the surface, our findings stand in contrast to previous work showing comprehenders to be more tolerant of ungrammaticalities or malapropisms in spoken L2 speech (e.g., Gibson et al., 2017; Lev-Ari & Keysar, 2012). These previous findings are consistent with the pattern that comprehenders would expect *more* errors in L2 speech and would process L2 speech *less* literally. One possibly relevant difference between this study and previous work is stimulus modality. We elected to present our stimuli in a written form so as to directly match the materials for each speaker (the written string was identical; what differed was only the perceived identity of the talker). Presenting spoken materials potentially introduces variability due to speaker intelligibility: individuals with non-Standardised accents may be harder to understand. This intelligibility difference might serve to heighten the overall perceived "noise" for an L2 speaker, leading to increased rates of non-literal interpretations to auditory stimuli; this difference is necessarily related to increased "noise" in comprehension and increased levels of comprehender difficulty.

The difference between comprehender-centred and producer-centred inferences about sources of "noise" connects with another important contrast between the present results and previous ones. Subject–verb agreement is a dependency between two elements, in contrast to lexical errors or preposition deletions, which reflect the mis-use of a single element. When faced with a mismatch between a subject and a verb, a comprehender could infer that either the noun or the verb is erroneous. Repairing the noun involves adding an *-s* to the head to make it plural, leading to a non-literal response. This non-literal response is based primarily on the reader's miscomprehension, as it is more likely for a reader to miss an *-s* than for a speaker to fail to add one. In contrast, one could also repair the utterance by assuming the verb was misproduced, reflecting a speech error. This production-centred inference would lead to a literal interpretation of the sentence's head.

As such, our results are consistent with the notion that readers are more likely to attribute an anomaly to their own miscomprehension if the ostensible speaker is more Standardised and more likely to attribute an anomaly to a speaker's misproduction if the ostensible speaker is from a different language group. This is compatible with both Gibson et al. (2017) and Lev-Ari and Keysar (2012). Assessing subject–verb agreement highlights the interplay between comprehender and producer-centred inferences about speakers, making testable predictions on both sides of the perception-production equation. We hope this sets the stage for future work with a similar eye towards dissociating the locus of "noise" in the system.

However, while the ostensible speaker mattered for agreement anomalies, there was no evidence of speaker-specific processing of the without-blends. This suggests that for an infrequently observed type of utterance that does not rely on recasting the utterance's grammatical form, assumptions about the speaker do not impact interpretation. There are several possibilities for this null result. One possibility is that readers did not have enough experience with utterances of this type to derive speaker-specific predictions. While attested, these utterances are rare, and comprehenders may not have sufficient experience to make speaker-specific predictions about the utterance's intended meaning. Another possibility relates to the semantic nature of the anomaly. Non-literal processing of without-blends might rely on semantic biases or processing heuristics, which may not be expected to differ by speaker group. This suggests the importance of general cognitive mechanisms for non-literal interpretations, as outlined in the good-enough processing framework (e.g., Ferreira, 2003). The first possibility might be more consistent with the noisy channel framework, while the latter might be more consistent with good-enough processing; both are reasonable explanations for the data.

Despite the lack of a speaker-specific effect in the without-blend items, comprehender experience did clearly matter for both types of items in both experiments. Experimental block was the single best predictor of performance in the present studies, such that more non-literal interpretations occurred in the beginning than the end of both experiments for both sentence types. This is consistent with adaptation on a short time-scale, such that readers accommodated to anomalies through the exposure provided to them in the experiment. It is also consistent with the influence of metalinguistic judgments on reading, such that subject–verb agreement and implicit negation were both made more salient by the nature of the comprehension questions. From this we conclude that at least small-scale experience is critical in shaping non-literal responses.[5] This is the core of both non-literal processing frameworks: past experience with language and the world influences present performance.

As such, the present data underscore the interconnectedness between what is perceived, what is understood, and what is learned, consistent with psycholinguistic frameworks that attempt to link production and comprehension mechanisms (e.g., Dell & Chang, 2014; MacDonald, 2013). Inferences rely on what meaning is expected of an utterance; these expectations can override the utterance's literal form. We have shown non-literal inferences to reflect inferences about a reader's miscomprehension and a speaker's misproduction. Taken in combination with previous work, our results suggest that both sources of "noise" matter for non-literal processing. This reflects the importance of considering comprehension in light of what is produced and production in light of what is likely to be comprehended.

Conclusion

Language occurs in context; as efficient computational machines, individuals are skilled at drawing upon their knowledge of regularities in language, indexical properties of other speakers, and the speaker's assumed communicative intent. This allows comprehenders to understand not just what has literally been said, but what was likely intended. This work shows that as readers, we show remarkable ability to shape our understanding in a rational and useful manner by drawing on properties of the language as a whole and on certain aspects of specific speakers.

Acknowledgements

The authors thank Neil Shook and Marta Millar for their assistance with stimuli development and data coding. They also thanks Frances Blanchette for her helpful discussions about implicit negation.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

Supported by a grant from the National Science Foundation (NSF OISE 0968369; PI: J. F. Kroll; co-PIs: P. E. Dussias and J. G. van Hell).

Supplementary material

The Supplementary Material is available at qjep.sagepub.com

Notes

1. In comprehension studies of subject–verb agreement, a preverbal adverb is often used to separate the local noun from the verb (see, for example, Wagers et al., 2009 for arguments). We elected to use stimuli containing a preverbal adverb to match this previous work, so as to make our items comparable to those used in existing literature in agreement processing. We selected six adverbs that were attested before a past tense copula (was/were) in the Corpus of Contemporary American English (COCA) corpus (Davies, 2008). Proportions of preverbal tokens are as follows: *actually*: 13%, *finally*: 8%, *literally*: 9%, *really*: 24%, *totally*: 1%, *probably*: 14%.
2. G*Power uses an effect size *f*, appropriate for a within-subjects repeated measures analysis of variance (ANOVA). We used this as an estimate for power in our studies despite using mixed-effect models for analysis as the two types of analyses are conceptually similar and as G*Power is a freely available, easy to use power calculator.
3. An alternate analysis is also possible, with the odds of "yes" versus "no" responses as a dependent measure. This means that one does not need to a priori decide which response is "non-literal" but makes interpretation less transparent. We have elected to use odds of non-literal responses upon suggestion of a reviewer for ease of exposition.
4. To add variability in item meaning and structure, we used some multiple-token ($n = 16$) and some single-token ($n = 32$) subject–verb agreement items. Excluding the multiple-token (distributive) items showed a similar numerical pattern in both experiments. The only statistical difference was that the three-way interaction in Experiment 1 was no longer significant ($p = .09$), as one might expect with a reduced sample size.
5. We ran additional analyses to see if any demographic information relating to participant language experience (education level, exposure to L2/regional dialect speakers) impacted non-literal response rates. These revealed no significant differences, which may be due to low variability in the data sample (as seen in Table 1); as such, we believe this warrants future research.

References

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., & Dai, B. (2014). *lme4: Linear mixed-effects models using Eigen and S4* (Version 1.1-7). Retrieved from https://cran.r-project.org/web/packages/lme4/index.html

Blanchette, F. (2013). Negative concord in English. *Linguistic Variation*, *13*, 1–47.

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, *23*, 45–93.

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). Native "um"s elicit prediction of low-frequency referents, but non-native "um"s do not. *Journal of Memory and Language*, *75*, 104–116.

Bosque, I., & Demonte, V. (1999). *Gramática descriptiva de la lengua española* [Descriptive grammar of the Spanish language]. Madrid, Spain: Colección Nebrija y Bello, Espasa.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*, 707–729.

Brehm, L., & Bock, K. (2013). What counts in grammatical number agreement? *Cognition*, *128*, 149–169.

Brehm, L., & Bock, K. (2017). Referential and lexical forces in number agreement. *Language, Cognition and Neuroscience*, *32*, 129–146.

Butterworth, B. (1982). Speech errors: Old data in search of new theories. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 73–109). Amsterdam, The Netherlands: Walter de Gruyter.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*, 368–407.

Christianson, K., Luke, S. G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 538–544.

Cutler, A. (1982). The reliability of speech error data. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 7–28). Amsterdam, The Netherlands: Walter de Gruyter.

Cutting, J. C., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory & Cognition*, *25*, 57–71.

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present*. Retrieved from https://corpus.byu.edu/coca/

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120394.

Drummond, A. (2013). *Ibex farm*. Retrieved from http://spellout.net/ibexfarm

Evans, J. S. B. T., Clibbens, J., & Rood, B. (1996). The role of implicit and explicit negation in conditional reasoning bias. *Journal of Memory and Language*, *35*, 392–409.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*, 164–203.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, *8*(10), e77661.

Foote, R. (2010). Age of acquisition and proficiency as factors in language production: Agreement in bilinguals. *Bilingualism: Language and Cognition*, *13*, 99–118.

Fraundorf, S. H., & Jaeger, T. F. (2016). Readers generalize adaptation to newly-encountered dialectal structures to other unfamiliar structures. *Journal of Memory and Language*, *91*, 28–58.

Frazier, L., & Clifton, C., Jr. (2011). Quantifiers undone: Reversing predictable speech errors in comprehension. *Language*, *87*, 158–171.

Frazier, L., & Clifton, C., Jr. (2015). Without his shirt off he saved the child from almost drowning: Interpreting an uncertain input. *Language, Cognition and Neuroscience*, *30*, 635–647.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, *47*, 127–152.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 8051–8056.

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*, 703–712.

Grey, S., & van Hell, J. G. (2017). Foreign-accented speaker identity affects neural correlates of language comprehension. *Journal of Neurolinguistics*, *42*, 93–108.

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*, 878–887.

Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, *32*, 277–307.

Horn, L. (1989). *A natural history of negation*. Chicago, IL: The University of Chicago Press.

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*, 30–66.

Ivanova, I., Branigan, H., McLean, J., Costa, A., & Pickering, M. (2017). Do you what I say? People reconstruct the syntax of anomalous utterances. *Language, Cognition and Neuroscience*, *32*, 175–189. doi:10.1080/23273798.2016.1236976

Jackson, C. N., Mormer, E., & Brehm, L. (in press). The production of subject-verb agreement among Swedish and Chinese second language learners of English. *Studies in Second Language Acquisition*.

Kaschak, M. P. (2006). What this construction needs is generalized. *Memory & Cognition*, *34*, 368–379.

Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General*, *133*, 450–467.

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review*, *64*, 459–489.

Labov, W. (1963). The social motivation of a sound change. *Word*, *19*, 273–309.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology*, *44*, 701–710.

Lev-Ari, S., & Keysar, B. (2012). Less-detailed representation of non-native language: Why non-native speakers' stories seem more vague. *Discourse Processes*, *49*, 523–538.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126–1177.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 21086–21090.

Lorimor, H., Jackson, C. N., & Foote, R. (2015). How gender affects number: Cue-based retrieval in agreement production. *Language, Cognition and Neuroscience*, *30*, 947–954.

Lorimor, H., Jackson, C. N., Spalek, K., & van Hell, J. G. (2016). The impact of notional number and grammatical gender on agreement with conjoined noun phrases in Dutch. *Language, Cognition and Neuroscience*, *31*, 646–661.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 226.

Nicol, J., & Greth, D. (2003). Production of subject-verb agreement in Spanish as a second language. *Experimental Psychology*, *50*, 196–203.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Attention, Perception, & Psychophysics*, *60*, 355–376.

Patson, N. D., & Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *The Quarterly Journal of Experimental Psychology*, *69*, 950–971.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*, 427–456.

R Core Team. (2014). *R: A language and environment for statistical computing* (Version 3.1.2). Vienna, Austria: R Foundation for Statistical Computing. Available from http://www.R-project.org

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition*, *36*, 201–216.

Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, *24*, 135–178.

Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language*, *76*, 195–215.

Tortora, C., & den Dikken, M. (2010). Subject agreement variation: Support for the configurational approach. *Lingua*, *120*, 1089–1108.

Trenkic, D., Mirkovic, J., & Altmann, G. T. (2014). Real-time grammar processing by native and non- native speakers: Constructions unique to the second language. *Bilingualism: Language and Cognition*, *17*, 237–257.

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, *38*, 584–589.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*, 206–237.

White, L. (2003). Fossilization in steady state L2 grammars: Persistent problems with inflectional morphology. *Bilingualism*, *6*, 129–141.