Published in final edited form as:

Lang Cogn Neurosci. 2015 July 1; 30(6): 635–647. doi:10.1080/23273798.2014.995109.

Without his shirt off he saved the child from almost drowning: interpreting an uncertain input

Lyn Frazier^a and Charles Clifton Jr.^b

^aDepartment of Linguistics, University of Massachusetts Amherst, Amherst MA, USA

^bDepartment of Psychology, University of Massachusetts Amherst, Amherst MA, USA

Abstract

Unedited speech and writing often contains errors, e.g., the blending of alternative ways of expressing a message. As a result comprehenders are faced with decisions about what the speaker may have intended, which may not be the same as the grammatically-licensed compositional interpretation of what was said. Two experiments investigated the comprehension of inputs that may have resulted from blending two syntactic forms. The results of the experiments suggest that readers and listeners tend to repair such utterances, restoring them to the presumed intended structure, and they assign the interpretation of the corrected utterance. Utterances that are repaired are expected to also be acceptable when they are easy to diagnose/repair and they are “familiar”, i.e., they correspond to natural speech errors. The results of the experiments established a continuum ranging from outright linguistic illusions with no indication that listeners and readers detected the error (the inclusion of *almost* in A passerby rescued a child from almost being run over by a bus.), to a majority of unblended interpretations for doubled quantifier sentences (*Many students often turn in their assignments late*) to only a third undoubled implicit negation (I just like the way the president looks without his shirt off.) The repair or speech error reversal account offered here is contrasted with the noisy channel approach (Gibson et al., 2013) and the good enough processing approach (Ferreira et al., 2002).

Keywords

syntactic blends; speech error reversal; noisy channel; syntactic processing; acceptability judgments; linguistic illusions

Various types of evidence suggest that perceivers can recognize dysfluencies (e.g., Clark & Fox Tree, 2002, Bailey and Ferreira, 2007) and some speech errors (e.g., Nooteboom & Quené, 2013). When speakers make a speech error, the errors are structured, not random (Coppock, 2006). In what follows we present the results of two experiments on the acceptability (Experiment 1) and acceptability plus interpretation (Experiment 2) of ungrammatical sentences modeled on attested speech errors involving syntactic blends. We

Corresponding author: Lyn Frazier, lyn@linguist.umass.edu.

Authors' addresses: Lyn Frazier, Department of Linguistics, University of Massachusetts Amherst, Amherst, MA 01003 USA, lyn@linguist.umass.edu, (413) 545 6827; Charles Clifton Jr., Department of Psychology, University of Massachusetts Amherst, Amherst, MA 01003 USA, cec@psych.umass.edu, (413) 545 2653

draw out the implications of the results for a theory of language comprehension and the relation between the grammar and judgments of acceptability.

How common are speech errors? Estimates vary, and in fact the notion of a “speech error” becomes quite problematic on the present view. To take a simple case,

consider (1). This attested example sounds like something a competent speaker of English might say using the grammar of English and trying to convey a particular meaning. On the other hand, the grammar of English only allows the affix *-ish* to attach to adjectives, not nouns, as illustrated in (2).

(1) “You want them all to be roughly the same size --ish”, spoken with a clear pause before “ish”. (The chef on the tv cooking show “Simply Ming” 1/2/14)

(2) *It’s roomish/computerish/measure-ish/woodish (vs wooden-ish).

Consequently, the status of an example like (1) is an issue: it may be exactly what the speaker intended to say, and thus not a speech error in one sense, but it may not conform to the general rules of the language, and thus be an error in another sense.

How such cases should be modeled is an interesting and complex question (see too Bever, Carroll & Hurtig, 1976, Bever, 2013, Frazier, in press, among others).

The central question about naturally occurring errors such as syntactic blends is not how frequent they are but whether the errors are systematic when they do occur. Coppock (2006) presents the most extensive investigation to date. She shows that her alignment model correctly distinguishes occurring from non-occurring syntactic blends. An actually occurring blend such as (3) tends to map words of the same syntactic category onto the same position.

(3) cast into question

- a. call into question
- b. cast into doubt

We take Coppock’s results to support the notion that some types of errors are natural in the sense that they accord with the principles of the sentence production system whether the errors have a high token frequency or not. In the present paper the focus will be on syntactic blends. Frazier and Clifton (2011a) reported blends involving a determiner quantifier and an adverbial quantifier. Rather than choosing between the two (*Many people thought...* and *People often thought...*), the actual utterances incorporate both of the options in a blend, as in (4).

(4)

- a. “Many people often thought that you use whipped cream pie” (National Public Radio, discussion of clowns and pie throwing)
- b. “...and it might not require scientific research to infer that the majority of sarcasm one encounters is usually spoken.” (UMass undergraduate paper)

- c. “A lot of people often think I’m insane when I tell them what I’m selling” (Fire cider)-NEPR interview 2/7/14

Frazier and Clifton (2011a) investigated the interpretation of doubled quantifier sentences containing *many-often*, *every-always*, *few-seldom*, and doubled explicit negation of various sorts. Speakers chose from interpretations containing a single quantificational force, the “undoubled” interpretation, and grammatical interpretations where each quantifier adds its own quantificational force, as illustrated in (5).

(5) Many students often turn in their assignments late.

What did that mean?

- A. The number of students who turn in their assignments late is large.
 B. The number of students who frequently turn in their assignments late is large.

Apart from the negation cases, the majority of responses indicated the interpretation where readers undoubled the quantification, that is, the interpretation that did not respect the grammatical form of the sentence. The exception was doubled negation, where the quantification (negation) was undoubled only half as often as for the other quantifiers.

The results are interesting in two respects. First, the results suggest that participants are repairing an utterance to the presumed intended form before interpreting it. Understanding the circumstances under which they do this is important for explaining the nature of the language comprehension mechanism, and its relation to the grammar. Second, it is interesting that doubled negation behaved differently. We suspect that this is because in the US, speakers of Mainstream American English are taught by their parents and in school not to use double negation.

Two theoretical claims of current interest, the “good-enough processing” proposal (Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007), and the “noisy channel” proposal (Gibson, Bergen, & Piantadosi, 2013; Piantadosi, Tily, & Gibson, 2012) also deal with interpretations that go beyond the grammar. However, they posit different kinds of explanations for such observations. While we accept at least some of the claims the positions make, we believe that it is possible to arrive at a deeper and more illuminating account of the observations that motivate them.

Before developing this claim, we present some experimental observations that we suggest will motivate it. Existing data on undoubling doubled quantifiers comes exclusively from off-line studies requiring participants to choose between two interpretations of a sentence (Frazier & Clifton, 2011a). This raises the issue of whether error reversal is primarily a matter of correcting written input under circumstances where deliberation is possible. This type of error reversal would still be of some interest, but of much more limited interest than an error reversal system that operates in both the visual and auditory modality under circumstances where deliberation is discouraged. To our minds, the most interesting outcome is if the speech error reversal system is a general, possibly automatic, part of language comprehension.

Experiment 1

To find out how general the speech error reversal system is, Experiment 1 tested the quantifier doubling sentences of Frazier and Clifton (2011a) in both the auditory and the visual modality, using a speeded acceptability judgment technique. In addition, Experiment 1 investigated “blend” examples where we suspect that reversing the speech error may well count as an illusion. These cases were originally pointed out to us by Greg Carlson, who noted the newspaper headline in (6).

(6) Mother saves child from nearly drowning.

Examples like (6) involve a blend of something like *prevent the child from drowning* and *the child almost drowned*. What is striking about (6) is that it definitely is not interpreted as the mother drowns her child, as would be required by the compositional interpretation.

Experiment 1A was conducted as an auditory study, and Experiment 1B as a nearly identical reading experiment. Because the interest was in whether speech error reversal is limited to circumstances encouraging deliberation, such as a choice between two interpretations, the task was a speeded acceptability judgment task. Also, we were interested in whether doubling examples might be judged good more often in an informal register than in a formal one. So each version of Experiment 1 was run with two between-participant conditions, one with fillers marked for an informal register and one with fillers appropriate in a more formal register such as standard academic speech. The question is whether participants might be more likely to accept blends in an informal register than in a formal one.

Methods

Materials—In each of Experiments 1A and 1B, we tested 16 “doubled” items (7) and 8 “blends” (“prevent near culmination”) (8) (see also (6)). The former items appeared in three forms, one in which the quantifiers were doubled (7a), and two in each of which a single quantifier appeared (7b) and (7c). The latter items contained the verb *save*, *prevent* or *keep*, and the modifier *almost* or *nearly*, and appeared in two forms, one with both the verb and the modifier (8a) and one with just the verb. Only 8 such items were included, because we wanted to avoid substantial repetition of the verbs and modifiers, and because we expected their acceptance frequency to be uniformly near ceiling, eliminating the need for more items.

(7)

- a. (doubled) Many students often turn in their assignments late.
- b. (control 1) Students often turn in their assignments late.
- c. (control 2) Many students turn in their assignments late.

(8)

- a. (blended) A passerby rescued a child from almost being run over by a bus.
- b. (nonblended) A passerby rescued a child from being run over by a bus.

All experimental items appear in Appendix 1.

These items were combined with 65 register-inducing items as illustrated in (9a) (informal) and (9b) (formal). Twenty of these fillers had various grammatical or semantic anomalies. In addition, the experiments contained 36 items investigating verb phrase ellipsis, and Experiment 1A contained 16 items investigating the effect of position of a prosodic boundary on the acceptability of some three-clause sentences.

(9)

- a. US doctors gotta go into debt to pay for their education.
- b. In the US, a doctor must pay for his or her education.

The materials for Experiment 1A were recorded in a sound-deadened chamber by a skilled and phonologically-trained linguist, attempting to pronounce them in a natural and not emphatically-accented fashion. The informal filler items were recorded with more casual phonology while the formal filler items and all the experimental items were recorded with more formal phonology, especially with respect to the release of word final stops. They were digitized as .wav files to be played to experimental participants.

Participants and procedures—Sixty University of Massachusetts undergraduates, who received course credit for participation, served in Experiment 1A. Thirty heard the list that contained the informal fillers (9a), and 30 the list with the formal fillers (9b). Six versions of each list were constructed to permit full counterbalancing of both types of items. Each item was tested equally often in each of its conditions, and each subject saw an equal number of items in each the conditions. Sixty different University of Massachusetts undergraduates participated in Experiment 1B, reading the same sentences (counterbalanced the same way) the Experiment 1A participants heard (30 with informal and 30 with formal fillers).

In Experiment 1A, participants were instructed that they were going to hear a series of sentences. As soon as a sentence completed, a fixation cross that was on during the sentence disappeared, and subjects could make an acceptability judgment. They were to quickly judge whether or not each was acceptable, and indicate their response by pressing a key on a computer keyboard (f for “unacceptable”, j for “acceptable”). The time from the end of the sentence to making the response was recorded by the computer. The experimental and filler sentences were played, in individually-randomized orders, over speakers in a sound-deadened chamber, following a series of 7 practice sentences in the formal or the informal register, as appropriate. Experiment 1B proceeded in the same way, except that a computer visually presented each sentence as a whole on a terminal. The experimental subject was instructed to read the sentence and as soon as possible, to press the f or j key to indicate a judgment of “unacceptable” or “acceptable.” The computer recorded participants’ responses and times to read and respond to the sentence.

Results

Listeners in Experiment 1A, and readers in 1B, were sensitive to the manipulation. The responses to the filler sentences that were used to induce the register difference differed depending on their formality: readers and listeners were less likely to accept informal register sentences than formal register sentences. In the auditory experiment, 78% of the

informal register-inducing sentences that were designed to be acceptable were accepted, while 89% of the corresponding formal sentences were. In contrast, 20% of the informal sentences that were intended to be anomalous were accepted, as opposed to 7% of the corresponding formal sentences. In the reading experiment, Experiment 1B, the corresponding values were 71%, 83%, 12%, and 14%.

Turning to the experimental items, the acceptability judgment results for the doubled items (8) for Experiments 1A and 1B appear in Table 1, both pooled over and separated by whether the quantifier was non-negative or negative (items 9-12 in Appendix A). A logistic mixed model analysis of the Experiment 1A data was conducted, using R version 2.15.1 (R Development Core Team, 2012) with the lme4 package version 0.999999-0. As was the case for all analyses reported in this paper, both random slopes and intercepts were used of conditions for participants and items. The analysis of Experiment 1 contrasted the pooled Control conditions with the Doubled condition (containing both negative and non-negative items) and the two Control conditions with each other. The only effect that was significant (taking a value of $t \geq 2.0$ to indicate an acceptable level of significance) was the difference between the Doubled condition and the pooled Control Conditions ($b = .78$, $SE = 0.21$, $z = 3.81$, $p < .001$). A similar analysis was performed on the Experiment 1B data. This analysis indicated that the Doubled condition was again accepted less often than the pooled Control conditions ($b = .80$, $SE = .36$, $z = 2.23$, $p < .03$), but this time the two Control conditions differed from each other ($b = .45$, $SE = .21$, $z = 2.13$, $p < .04$). Further, in Experiment 1B, acceptance was higher in the Informal than the Formal register condition ($b = .86$, $SE = .36$, $z = 2.39$, $p < .02$). No interaction approached significance ($z < 1$). As can be seen in Table 1, the negative items differ greatly from the non-negative items, presumably because our participants were instructed as children to avoid double negatives.

The results for the blend items (8) for the two experiments appear in Table 2. No statistical analyses are necessary for the data in Experiment 1A. The central result is that participants were as likely to accept a prevent-near-culmination blend as its nonblended, completely plausible, counterpart, and both were comparable in acceptance to the controls for the doubled items. However, to assess the apparently-lower acceptance rate for Blended items in the Formal condition of Experiment 1B, an analysis like that used for the Doubled items was conducted. While the difference between the Formal and Informal register conditions, evaluated for Blended items, was significant ($b = 1.70$, $SE = 0.8$, $z = 2.09$, $p < .04$), the interaction between register and Blended vs Nonblended was not significant ($b = 1.02$, $SE = 1.25$, $z = 0.81$, $p > .41$).

The times to make acceptability judgments about the doubled quantifier sentences appear in Table 3. Little can be made of the absolute time values, since they were measured from the offset of the sentence in Experiment 1A, and the visual onset of the sentence in Experiment 1B. A linear mixed model analysis of the response times in Experiment 1A indicated that, overall, formal sentences were responded to more rapidly than informal sentences ($b = 79.3$, $SE = 26.3$, $t = 3.01$) and the doubled condition was responded to more slowly than the pooled control conditions ($b = 30.8$, $SE = 15.4$, $t = 2.00$). No other effects or interactions approached significance. The pattern of results for Experiment 1B was similar, except that the difference between registers was nonsignificant ($t < 1$), while responses to doubled items

were significantly slower than responses to the pooled controls ($b = 148$, $SE = 55$, $t = 2.68$). No other effects were significant.

The previous analysis included all responses, both “acceptable” and “unacceptable.” Because it is possible that judgments of “unacceptable” will generally be slower than judgments of acceptable, a second analysis was conducted on acceptable (unblended) responses only. The pattern of response times was very similar to the pattern for all responses. In Experiment 1A, the mean “acceptable” RT to doubled sentences was 476 ms and to the undoubled control sentences, 426 (a difference that did not reach significance; $b = -27.97$, $se = 16.58$, $t = 1.69$). In Experiment 1B, the corresponding means were 3075 and 2568, a difference that was fully significant ($b = 166$, $SE = 53.9$, $t = 3.077$). Taken together, we conclude that responses to doubled sentences are slower than to controls, even when they are judged acceptable.

The mean response time for the blend sentences (items like (9)) was 368 ms in Experiment 1A, and 3173 ms in Experiment 1B. No effect involving either register or blending approached significance in either experiment.

Discussion

Experiment 1 shows that some blends achieve the status of true illusions. Apparently participants did not even notice the oddity of the-prevent near-culmination sentences since they were accepted as often as their nonblended counterparts. What makes them so convincing? At present we can only speculate. It may be that the conceptual predictability of preventing or saving/keeping someone from a bad event allows the listener to grasp that an Agent prevented a bad event, and the bad event almost happened. In effect, the predictability of the undoubled meaning and the “silliness” of the literal meaning may conspire to make the listener/reader confident about the intended meaning of the sentence. This fits with informal attempts to get judgments about these sentences from our friends and relatives: non-linguists just don’t see anything wrong with them. Academics, once pressed, see the problem. This casual intuition of the sentences being entirely unobjectionable is of real interest, we think.

Linguistic illusions are interesting. The prevent-near-culmination illusion may be similar to the classic illusions like: If a plane went down exactly on the border between the US and Canada, where would you bury the survivors? Attention is not directed to the erroneous part of the sentence, and the message has already been extracted before the error is encountered.

Turning to the doubled quantifiers, it is clear that they are not terrible sentences, since they are often accepted, if not at rates comparable to the nonblended counterparts. We think the relatively high rate of acceptance is not surprising since doubled quantification does not seem unnatural, and we expect high acceptability ratings in general for grammatical utterances, for familiar utterances, and for natural utterances.

The slowdown for the doubled quantifier sentences, fully significant for both written and spoken materials and for written materials when only “acceptable” responses were analyzed (and marginally significant for spoken materials in this latter case), fits with the proposal

that the input is syntactically restored to what the reader or listener presumes is the intended form/interpretation, taking some time to do the syntactic restoration. The absence of a slowdown for the blended sentences fits with the other evidence indicating that readers and listeners simply do not detect the doubling in the prevent-near-culmination sentences, and it casts doubt on a possible proposal that the slowdown in the doubled quantifier examples is a simple artifact of the extra word since in that case it should also have been observed for the prevent-near-culmination sentences but not in the spoken sentences, where responses were timed from the sentence offset.

Experiment 1 indicates that speech error reversal is not limited to a deliberative process that occurs during the processing of written, possibly edited, language. In Experiment 1, doubled quantifier sentences were accepted three quarters of the time in auditory language processing involving normal presentation rates and speeded acceptability judgments. This suggests that the error reversal process is a part of the general comprehension mechanism. Further, the prevent-near-culmination sentences also show massive unblending as indicated by acceptance rates at near ceiling.

Turning to the register manipulation, the sentences inducing a particular register were accepted more often in the formal register than the informal register. A trend in the opposite direction appeared for blended items, where acceptance was numerically higher in the informal than in the formal register (considering only the doubled items, 79% vs 71% in Experiment 1A, 81% vs 72% in Experiment 1B), but was significant only in Experiment 1B. This tendency to be more lenient in the informal register did not interact with doubled vs controls, so we cannot claim that it reflects anything more than an overall bias toward acceptance. Further, response times were slower in the informal than the formal register for auditory sentences. Examination of Table 3 suggests that this effect is mostly due to the slow times for doubled sentences in the informal register, which could reflect some conflict between noting the oddity of the doubled sentence and the willingness to accept this oddity in the informal condition. However, the interaction that would support this suggestion was not significant. Overall, there is little reason to conclude that the difference between formal and informal register had any interesting impact on how listeners and readers processed and evaluated sentences that could reflect speech errors.

In Experiment 1 only acceptability judgments were collected. In the case of doubled quantifier sentences we have data from Frazier and Clifton (2011a) about interpretations. In the case of prevent-near-culmination sentences, intuitions strongly support the assumption that interpretations were based on the repaired/unblended input since the literal interpretation is highly implausible and thus would be expected to lower ratings. In Experiment 2 both interpretations and ratings were collected in an effort to assess this intuition and to examine the relation between interpretations and ratings for the doubled quantifier sentences.

Experiment 2

The second experiment collected either an interpretation or a naturalness rating from each participant for each of the 16 doubled quantification sentences and the 8 prevent-near-

culmination “blend” sentences used in Experiment 1. It collected a rating and an interpretation for each item to assess the possibility that, for at least some items, the rated acceptability would co-vary with the frequency with which the items were taken literally vs. being repaired. Each subject gave only one response (rating or interpretation) to each item because one judgment could interfere with the other one. For instance, being forced to make an interpretation might direct a reader’s attention to the grammatically unlicensed relation between form and interpretation, and affect the rating. Since judgments were counterbalanced over items across subjects, we thus obtained unbiased item-by-item measures of interpretation and acceptability.

In addition, Experiment 2 explored a novel structure. Consider (10), a naturally occurring syntactic blend in (10a), and its nonblended and presumably intended version in (10b) (cf. (10c), the presumed other source of the blend in (10a)). Experiment 2 addressed the question of whether participants will interpret the utterance as un-blended

(10)

- a. I just like the way the president looks without his shirt off.
- b. I just like the way the president looks with his shirt off.
- c. I just like the way the president looks without his shirt.

Possible interpretations:

- 1. According to me, the president looks good with his shirt off.
- 2. According to me, the president looks good with his shirt on.

Methods

Materials—The 16 doubled-quantifier and the 8 prevent-near-culmination sentences used in Experiment 1 were used in Experiment 2. In addition, a 7-point naturalness rating scale (1 = very unnatural, 7 = very natural) and a 2-choice question (as used in Frazier & Clifton, 2011a) was constructed for each one. An example of the latter appears in (4), earlier. In addition, 12 “without × off” sentences like the one illustrated in (10a) and (10b) were constructed, and a naturalness scale and 2-choice interpretation question was written for each of them. Two versions of each of the latter two types of sentences were written, one with the blend and one without the blend as illustrated in (10b). Only a single version of the doubled-quantifier sentence was used in Experiment 2 (i.e., no non-blended control sentences were presented). All experimental sentences appear in Appendix 2.

In addition to these 36 experimental sentences, 84 experimental and filler sentences of a variety of forms (including sentences with relative clauses, verb phrase ellipses, and temporary closure ambiguities, some of which were confusing and arguably ungrammatical) were constructed. Each of these had its naturalness scale and its two-choice interpretation question.

The resulting list of 120 items was divided into two forms. In each form, half of the items of each kind were followed by a 7-point naturalness rating scale, and the other half by a two-

choice interpretation question. Items were counterbalanced across forms, so that an item that was rated in one form was interpreted in the other form.

Participants and procedures—A total of 64 University of Massachusetts undergraduates were tested in individual half-hour sessions, 32 in each of the two forms of the list. They were instructed to read each sentence for understanding and press a button to indicate comprehension, and then read and answer the question that appeared after the sentence disappeared. They were told that they would receive different types of questions, sometimes about how natural the sentence seemed, and sometimes about what the sentence meant. After receiving three practice items, the 120 items in the experimental list were presented on a computer terminal in individually-randomized orders. A computer recorded their responses and reaction times.

Results

The core data of Experiment 2 are the ratings and proportions of undoubled/unblended interpretations for the three types of sentences. These values appear in Table 4.

Descriptively, even including the sentences with negative quantifiers (items 9-12 in Appendix 2), it is clear that the doubled quantifier sentences are considered to be moderately natural, and receive undoubled interpretations a small majority of the time.

Turning to the remaining sentences, it is very clear that participants “read thorough” the blending of the verb and the modifier of the “prevent-near-culmination” sentences, hardly noticing their clash. Neither the difference in ratings nor the difference in interpretations approached significance in mixed model analyses ($t < 1.0$). In contrast, concerning the “without his $\times$ off” sentences, participants quite consistently noted the blend of the negative *without* and the final modifier (e.g., *off*) and rated the items with such a blend consistently low. The difference in mean ratings was highly significant ($b = .96$, $SE = .16$, $t = 5.89$), as was the difference in logit proportion of undoubling ($b = 3.94$, $SE = 0.5$, $z = 7.75$, $p < .001$)

Probing more deeply into the data, one might ask whether the proportion of undoubled/unblended responses and their ratings correlated on an item-by-item basis. The simple answer is, not really, apart from a very few deviant cases. Mean (across subjects) naturalness ratings and proportion of undoubled interpretations were computed for each of the doubled-quantifier and the without-his-X-off sentences (there was inadequate variability in interpretations of the prevent-near-culmination sentences to support the correlational analyses reported below). Consider first the doubled-quantifier sentences. The Pearson product-moment correlation between proportion undoubled and naturalness rating was a moderately high 0.49, but inspection of Figure 1 indicates that the negative items mentioned above are almost entirely responsible for the observed correlation (particularly items 9 and 12).¹ Eliminating these four negative items reduces the correlation to a nonsignificant $r = 0.11$.

¹Examination of Figure 1 makes it clear that there were substantial differences among the four double-negative items. Items 9 and 12 are consistently interpreted as having double negation and rejected, following prescriptive grammar, but items 10 and 11 may reflect the recognition that some double negative forms are available in certain dialects - including, for item 11, the dialect of the current experimental participants.

Consider next the “without his $\times$ off” items. The overall correlation between proportion of unblended interpretations and ratings was .71. However, this large correlation is almost entirely due to the separation of the items that contained blends and items without blends, as can be seen in Figure 2. The correlation for the items with blends by themselves was only .08.

Discussion

The lowest rate of unblending occurred for “without his $\times$ off”, (34%), compared to doubled quantifiers (62%), and prevent-near-culmination items (95%). The acceptability penalty for the category with the most undoubling (prevent-near-culmination) was .2 compared to a penalty of .9 for the category with the least undoubling (“without his $\times$ off”). The lower rate of unblending for these latter sentences might be a matter of the plausibility of the literal interpretation. The speaker may really like the way the president looks with his shirt on, but not with it off. For the comprehender, it may be likely that the speaker made an error, but it isn’t certain. By contrast, for prevent-near-culmination, the comprehender can be quite certain about the intended meaning. However, the lack of correlation between rate of unblending and acceptability (except at the extremes) does pose some interpretive problems. Frequent identification of the need to unblend or undouble an item does not by itself result in low acceptability, nor do factors such as the plausibility of the blended (or the unblended) meaning seem to have the same effect on acceptability ratings and the probability of rejecting the literal syntax.

General discussion

We have presented evidence that a variety of sentences, modeled on natural speech errors, are interpreted in a non-literal fashion, as if the reader or listener repairs the errors that they represent. The sentences we have examined here and in prior research (Arregui et al., 2008; Frazier & Clifton, 2011b) differ systematically in how frequently they are repaired and how acceptable they are taken to be. We propose that these systematic differences can be understood in terms of what is known about sentence reanalysis (Fodor & Ferreira, 1998), on the one hand, which governs repairs, and how natural the pairing of utterance and form, i.e., how natural the speech error is, on the other hand. In terms of reanalysis, it has been shown that reanalysis costs are a function of the clarity of the error signal, the implausibility of the first analysis (easy reanalysis involves an implausible first analysis), the plausibility of the final/corrected analysis, and the number of repair operations needed. Without evidence that a repair is needed, no repair will be undertaken. Acceptability, we suggest, is a function of reanalysis costs plus the naturalness of the speech error being repaired: is the form the parser receives one that might have been produced by the human production system with the intent to express the meaning hypothesized.

The “good-enough” approach to processing (Ferreira, Bailey, & Ferraro, 2002; Ferreira & Patson, 2007) and the “noisy channel” approach (Gibson, Bergen & Piantadosi, 2013; Piantadosi, Tily, & Gibson, 2012)) claim that listeners and readers sometimes assign interpretations to sentences other than those licensed by the grammar (the syntax plus the compositional semantics). The “good-enough” approach has been described in a variety of

ways. It sometimes appears to claim that a language comprehender simply overlooks the requirements of grammar when motivated to do so, attending only to factors such as plausibility and frequency. Sometimes the approach is associated with claims that analyses are “underspecified,” but it is not clear just what grammatical properties are left unspecified (Ferreira & Patson, 2007). At still other times, it appears to suggest that reader or listeners will fall back on plausibility (etc.) when grammatical processing appears to be too difficult, or when they are having difficulty recovering from some kind of error of processing or interpretation (Christianson et al., 2001). Our proposal differs sharply from the first construal (although we acknowledge that someone who is not paying attention to what is read may fail to process it). Similarly, although carefully spelled-out notions of underspecification may help illuminate the nature of language processing (cf. Frazier & Clifton, 1996; Marcus, Hindle, & Fleck, 1983), we do not think that underspecification applies to the data we have presented here.

In contrast to the good-enough position, we do not claim that that readers and listeners interpret a sentence in a way that clashes with its grammatically-licensed interpretation as a result of their not putting enough effort into processing it - as a result, that is, of processing it in a good-enough fashion for the current demands. Rather, we claim that readers and listeners deploy an extensive and sophisticated knowledge of the types of errors speakers and writers are likely to make, and combine this knowledge with their understanding of language structure to reach an informed judgment of what the writer or speaker intended to convey. Natural blends are overlooked, and highly unlikely messages are disregarded - but just in case there is a plausible way in which a message could have been distorted in the course of being realized given the properties of the human production system. We believe that the observation in Experiment 1, that sentences with doubled quantifiers take longer to process than sentences with a single quantifier, is evidence for substantial and sophisticated processing, and argues against readers and listeners accepting a shallow, “good-enough,” interpretation.

One way to develop a Good Enough approach to doubled quantifier data, suggested to us by a reviewer, might be to note that readers don’t process very deeply semantic information that fits into context (consider, for instance, the “Moses” illusion; Erickson & Mattson, 1981; Sanford & Sturt, 2002), and thus don’t notice the addition of the second quantifier in doubled quantifier sentences because its meaning is so similar to the meaning of the first quantifier. One problem with this account is that there is substantial evidence that readers and listeners still process the input syntactically (Frazier, 2008b), but in the present research, subjects seem to treat doubled and blended sentences as if they had a different syntactic structure, e.g. one with only a single quantifier. Another problem with this sort of account is that one doesn’t generally ignore constituents just because they have similar meanings. As pointed out in Frazier and Clifton (2011a), sentences like (11) may receive an interpretation where I will go the day after tomorrow despite the fact that *tomorrow* generally is anchored rigidly to the speech time. At least some of the time, rather than taking the second occurrence to be redundant it is taken to be informative and interpreted with respect to a shifted time.

(11) Tomorrow I will go tomorrow.

An even more problematic example is the past perfect. Since the perfect already implies completion, the past should simply drop out because its meaning is so similar.

(12) Yesterday I had gone to the store at 6 so by 7 I could work on the manuscript. In (12), why don't perceivers simply take the past (expressed three times) to be redundant. Past perfects would never be perceived except when the task demanded a distinction. In sum, dropping out the meanings of words because they are similar to the meanings of other words in the sentence may happen on occasion. But any theory based on this approach owes an account of why we obey something like the full interpretation principle in other sentences.

One final problem for a good-enough approach appears in the effects of register. Experimental items were generally rated as more acceptable in the informal than in the formal conditions. A good-enough approach could suggest that this reflects shallower processing in the informal condition, resulting in a lowered criterion for acceptability. However, the filler items were rated as less acceptable in the informal than in the formal condition. This is inconsistent with an across-the-board adjustment in depth or extent of processing as suggested by the good-enough approach.

In the “noisy channel” approach, Gibson and colleagues (Gibson et al., 2013; Piantadosi et al., 2012) have noted that environmental noise, producer errors, and perceiver errors make the input to the language system less than error-free. The signal may be distorted in systematic and in random ways. Recovering a meaning from a noisy input may involve rational statistical inferencing. Gibson and colleagues suggest that a comprehender estimates the probability of the intended message (s_i) given what was perceived (s_p), i.e., $P(s_i | s_p)$. Following Bayes rule, this estimate is the product of the prior probability of s_i multiplied by the probability that this message would be realized as s_p , $P(s_p | s_i)$. Acceptance of a message is then governed by this posterior probability.

Our proposal is not inconsistent with the outline of the Gibson et al. proposal, but it spells out the cognitive mechanisms that may underlie s_i and $P(s_p | s_i)$. Gibson et al. make some suggestions about mechanism. They say, very reasonably, that more plausible messages will have a higher s_i than less plausible ones. They propose, equally reasonably, that $P(s_p | s_i)$ is higher when fewer changes have taken place during transmission than when more changes have occurred. Further, they propose a “Bayesian size principle” that makes it more likely that, as a message goes through the noisy channel, material will be deleted from an intended message than added to it. They say “a deletion only requires a particular word to be randomly selected from a sentence, whereas an insertion requires its selection from (a subset of) the producer's vocabulary” (Gibson et al., 2013, p 8053). According to this latter claim, a reader or listener is more likely to take a sentence that is missing some elements to express an intended message that requires these elements than to take a sentence with extra elements in it to express a message that does not require them.

We doubt that this prediction is quite correct. It does seem plausible when considering errors introduced in transmission from message to utterance. And Gibson et al. do provide some evidence for it. For instance, they showed that people were more likely to say TRUE to the plausible *Did the daughter receive something?* after reading *The mother gave the candle the*

daughter (a one-deletion sentence) than after reading *The mother gave the daughter to the candle* (a one-insertion sentence). However, some evidence suggests that this is not the entire story. Consider active vs. passive sentences. An active sentence has two deletions from a passive message (at least - the auxiliary and the *by*), while a passive sentence has two insertions. Gibson et al. found no differences in ungrammatical interpretation frequencies of actives and passives; both received nearly 100% grammatically-justified interpretations. However, Ferreira (2003) used sentences that were likely more strongly biased in terms of plausibility (e.g., *The man bit the dog*) than the Gibson et al. sentences, and asked listeners to explicitly indicate the agent or the patient. She found that passives were misinterpreted as actives more often than vice versa (e.g., an average of 79% accurate vs 95% accurate for her implausible reversible sentences in Experiment 1). This reflects a higher rate of misinterpretation of sentences with insertions than with deletions, the opposite of the Gibson et al. prediction.

The position we advance here makes some alternative suggestions. It focuses on comprehension processes that go beyond a kind of Bayesian estimation of the probability of a particular message, given transmission across a noisy channel. In common with Gibson et al., we propose that the likelihood that the speaker intended a particular message (in their terms, $P(s_i)$) does play a role. We go beyond Gibson et al. though in proposing that memory and interpretation processes that take place during comprehension have their effects, and crucially, we propose that a reader/listener's knowledge of what kinds of speech errors are natural has an effect. If a comprehender implicitly knows that the human production system might make a particular type of error, the comprehender will be more willing to overlook the error and "read through" to the intended message.

Consider active vs. passive sentences. It is known (Mehler, 1963) that passives are misremembered as actives more often than vice versa. Arregui et al. (2008) showed that "mismatching" VP ellipsis sentences in which the antecedent was a passive and the elliptical phrase was an active (e.g., *The dessert was praised by the customer after the critic did already*) were judged to be more acceptable than such sentences with an active antecedent and a passive ellipsis (*The customer praised the dessert after the appetizer was already*). This is not what Gibson's Bayesian size principle would predict. Rather, it follows from either of two possible processes. One possibility is that the reader/listener would have a tendency to misremember the passive antecedent as an active during processing, resulting in a memory match to the active ellipsis and thus an acceptable sentence. Alternatively, the reader/listener has implicit knowledge that the speaker/writer was likely to have misremembered the passive antecedent as active before producing the ellipsis, resulting in producing a VPE sentence with a mismatch that is understandable, and thus relatively acceptable.

Another example of how a possible memory error can affect interpretation appears in Frazier and Clifton (2011b), who examined the interpretation of VP ellipses like *George does too* following sentences like *Ian travels in winter and he stays home in summer*. The ellipsis was taken by experimental participants to mean that George travels in winter and stays home in summer (reflecting conjunction of VPs) just as often (85% of the time) when the antecedent contained conjoined clauses as when it contained conjoined VPs. This contrasts with a substantially lower frequency of such interpretations (55%) when the two clauses had

appeared as separate sentences.. A speaker or writer may produce an antecedent clause with conjoined clauses with coreferential subjects, forget that conjoined clauses were produced rather than conjoined VPs, and then incorrectly produce an ellipsis clause with an elided VP intending a conjoined VP antecedent

We propose that comprehenders take into account the plausibility of a message and the presence or absence of a speaker's motivation for producing a particular error. Consider (13), where there is pragmatic motivation for error in (13a), namely, not insulting one's interlocutor, but there is not pragmatic motivation for the error in (13b), where the speaker would be going out of her way to insult the interlocutor. In a rating study (Frazier, 2008a), (13a) was rated as better than (13b): on a scale of 1-5, where 5 is fully acceptable, (13a) received a rating of 4.05 whereas (13b) received a significantly lower 3.37. This asymmetry is not explained by the noisy channel approach as developed by Gibson et al. because there are an equal number of deletions and insertions in (13a) and (13b).

(13)

- a. If you think this is going to solve the terrible problems in Najaf, we're deluding ourselves. 4.05
- b. If we think this is going to solve the terrible problems in Najaf, you're deluding yourself. 3.37

Another example where an error motivated by the production system is rated higher than one that is not involves contexts where a surprising feature might, or might not, be motivated by anticipation of an upcoming feature. In unpublished studies with Jennie Mack, we looked at copy raising sentences like (14). The sentence in (14b) is less acceptable (being rated at 2.40 on a 5-point acceptability scale) than the sentence in (14a) (2.83). While no difference is predicted between (14a) and (14b) on the noisy channel view (both require the deletion of a single number feature), in (14a) the upstairs plural is motivated by an upcoming plural whereas in (14b) it is not.

(14)

- a. There seem like there are some problems with the project
- b. There seem like there is a problem with the project.

A noisy channel approach to comprehending input in a way that goes beyond the grammar does not have to be concerned with what goes into the prior or what underlies the probabilities linking form and intended message. By contrast, the present approach attempts to characterize systematic properties of the performance system which have consequences for the ways in which linguistic input is structured and interpreted. Systematic properties include, among others, the effects on comprehension of regularly having a choice between alternate syntactic formulations of the same message. Frazier and Clifton's (2011b) observation that VP ellipses were interpreted the same following conjoined VPs and following conjoined clauses, which was presented as illustrating the role of memory factors in interpreting mismatch ellipsis, provides an illustration of such systematicity. The systematic choice between formulating a sentence using conjoined VPs vs. conjoined

clauses (with coreferential subjects) could result in a conjoined clause antecedent for an elided VP, which actually requires a conjoined VP antecedent. Frazier and Clifton's (2011b) readers interpreted such a sentence in the same way as one with a conjoined VP antecedent, presumably reflecting their realization that a writer or speaker could have forgotten which alternative formulation had been chosen when it came time to formulate the elliptical clause. Understanding this type of mechanism relating production, comprehension and the grammar is central, we think, to understanding how the language system works. Simply relating intended message and probability of a particular interpretation does not reveal the kinds of mechanisms involved.

In more general terms, we view the good-enough position and the noisy channel hypothesis as providing motivation for a deeper exploration of how readers and listeners interpret sentences in ways that deviate from what a grammar-based compositionality dictates. We do not challenge the claim that readers and listeners sometimes arrive at interpretations that are not grammatically-justified. Neither do we question the value of the Bayesian analysis of the noisy channel hypothesis for beginning to break the sources of these "mis"interpretations into prior and conditional probabilities. Our approach has been to try to understand how people assign structure and interpretations during comprehension, and why they behave this way. In our sort of approach, it is not necessarily desirable to pack all information influencing expectation into one statistical inferencing model. For example, noise from potential speech errors might influence the processes underlying comprehension in different ways than does acoustic noise in the environment. According to the noisy channel perspective though, this may be viewed as a theory-internal question about whether both sources of noise contribute to a single rational inferencing model, or to distinct models for the two types of noise. If it should turn out that every aspect of sentence comprehension is expectation driven, the difference between these two types of noisy channel accounts (single model vs separate models) may not matter at least with respect to the pairing of an input and the probability of computing a particular type of interpretation. But it may still matter for a process model, where the two types of information might influence different phases of processing, might interact with different types of information, or might differ in terms of whether the noise is consciously available to the comprehender.

Conclusions

The reversal of speech errors is an interesting phenomenon. In studying this phenomenon, we have focused on syntactic blends (mismatch ellipsis, doubled quantification, prevent-near-culmination, and *without his × off*). The probability of reversing an error seems to be a function of the same properties that determine syntactic reanalysis in the processing of garden path sentences. The nature of the error signal matters. The clearer the error signal, the easier the revision/repair. The implausibility of the initial/literal analysis is one good signal. The ease of making the revision/repair in terms of the plausibility of the final analysis (see Frazier & Clifton, 1998) and the number of operations (Arregui et al., 2008) also seems to matter. In contrast to the Gibson et al. "Bayesian size principle" (according to which it is more likely that material will be deleted than added as it goes through the "noisy channel"), listeners and readers seem quite willing to accept that a speaker or writer could add redundant material in the course of expressing an intention. A comprehender could then

disregard (e.g.) doubled quantification rather than taking the multiple quantifiers literally. But diagnosing the need for a revision/repair can create difficulty. We suggest that is the difficulty in the *without his × off items* investigated in Experiment 2.

On the current approach, general likelihood doesn't fully deal with the heart of the matter; but assessing speakers' intentions and speakers' limitations does. Listeners and readers do have to cope with a noisy channel and they have various coping strategies, such as guessing or saying *what?* But with respect to inputs conforming to the constraints of the human production system, it looks as if listeners and readers are employing their implicit knowledge of the production system together with their knowledge of the grammar of a language to interpret deviant utterances.

Acknowledgments

This research was supported in part by grant HD18708 from NICHD to the University of Massachusetts. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The authors are also grateful to Ted Gibson's lab, especially Ted Gibson, Leon Bergen, and Richard Futrell for discussion of the noisy channel approach. References

Appendix 1. Materials used in Experiment 1

- A. Doubled items. All three versions – doubled and controls – are shown for item 1. Only the doubled versions are shown for the remaining items
1.   a. Many students often turn in their assignments late.
  - b. Students often turn in their assignments late.
  - c. Many students turn in their assignments late.
  2. Many shopkeepers often mark down prices.
  3. Many journalists often interview local authors.
  4. Many British gardeners often fertilize plants.
  5. Every teacher always repeats the main point of a class.
  6. Every customer always asks about a discount during summer shopping trips.
  7. Every Dutch tourist always speaks English beautifully.
  8. Every accountant always checks the social security number of dependents.
  9. No social studies teachers never go to see the principal.
  10. Friends don't hardly ever criticize each other.
  11. I miss not going to the gym.
  12. Jane doubts Bill didn't go to Paris.
  13. Few Americans seldom speak three languages.
  14. Few acquaintances seldom notice a slight social offense.
  15. Few politicians seldom directly criticize their opponents.

16. Few announcers seldom make noticeable speech errors.
- B.** Blend items. Both forms are shown for item 1. Only the blended item (with the quantificational adverb added) is shown for the remaining items.
1.
    - a. A mother saved her child from nearly drowning.
    - b. A mother saved her child from drowning.
  2. A stranger rescued a dog from almost being strangled by his leash.
  3. The neighbor prevented an older lady from almost dying of a heart attack.
  4. The babysitter kept the baby from nearly suffocating.
  5. The nurse saved a frail patient from nearly falling out of bed.
  6. A passerby rescued a child from almost being run over by a bus.
  7. Grandfather prevented a horse-driven carriage from nearly tipping over.
  8. The rescue squad kept the bomb from almost going off.

Appendix 2: Materials used in Experiment 2

- A.** Doubled items. The full paradigm (with both rating scale and interpretation question) is shown for item 1. Only the sentence is shown for the remaining items.
1. Many students often turn in their assignments late.

Q. How natural was that sentence? {very unnatural .. ... very natural}

Q. What did that mean? 1. The number of students who turn in their assignments late is large. 2. The number of students who repeatedly turn in their assignments late is large.
  2. Many shopkeepers often mark down prices.
  3. Many journalists often interview local authors.
  4. Many British gardeners often fertilize plants in late spring.
  5. Every teacher always repeats the main point of a class.
  6. Every customer always asks about a discount during summer shopping trips.
  7. Every Dutch tourist always speaks English beautifully.
  8. Every accountant always checks the social security number of dependents.
  9. No social studies teacher never goes to see the principal.
  10. Friends don't hardly ever criticize each other.
  11. I miss not going to the gym.
  12. Jane doubts Bill didn't go to Paris.
  13. Few Americans seldom speak three languages.

14. Few acquaintances seldom notice a slight social offense.
15. Few politicians seldom directly criticize their opponents.
16. Few announcers seldom make noticeable speech errors.

B. “Without his X” items. The full paradigm is given for item 1. Only the “without” (blended) version is given for the remaining items.

1.   **a.** I just like the way the president looks without his shirt off.
- b.** I just like the way the president looks with his shirt off.

Q. How natural was that sentence? {very unnatural .. ..
..... very natural}

Q. What did that mean? 1. I think the president looks good with his shirt off. 2. I think the president looks good with his shirt on.

2. Sam left without his coat off.
3. Sophie cleans the house without her clothes off.
4. Lulu hated the way she looked without her hat off.
5. Stan arrived without his luggage missing
6. Bart went out without his credit card at home.
7. Anita managed without her boss away.
8. Jena was depressed without the dog gone.
9. Tim thought Mary was sad without her spouse away.
10. The protestors had to leave without their supporters absent.
11. Tierney claimed Aiden pouted without his parents gone.
12. Jason suggested David fought back without the teacher absent.

C. Prevent near culmination sentences. The full paradigm is shown for item 1; only the sentence with the quantificational adverb is shown for the remaining items.

1.   **a.** A mother saved her child from nearly drowning.
- b.** A mother saved her child from drowning.

Q How natural was that sentence? {very unnatural .. ..
..... very natural}

Q What did that mean? 1. The child drowned. 2. The child didn't drown.

2. A stranger rescued a dog from almost being strangled by his leash.
3. The neighbor prevented an older lady from almost dying of a heart attack.
4. The babysitter kept the baby from nearly suffocating.

5. The nurse saved a frail patient from nearly falling out of bed.
6. A passerby rescued a child from almost being run over by a bus.
7. Grandfather prevented a horse-driven carriage from nearly tipping over.
8. The rescue squad kept the bomb from almost going off.

References

- Arregui A, Clifton C Jr, Frazier L, Moulton K. Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language*. 2006; 55(2):232–246. [PubMed: 17710192]
- Bailey, KGB.; Ferreira, F. The processing of filled pause disfluencies in the visual world. In: van Gompel, RPG.; Fischer, MH.; Murray, WS.; Hill, RL., editors. *Eye Movements: A Window on Mind and Brain*. Elsevier Inc.; Oxford, UK: 2007. p. 485-500.
- Bever, TG. The biolinguistics of language universals – the next years. In: Montserrat, Sanz; Laka, Itziar; Tanenhaus, Michael K., editors. *Language Down the Garden Path. The Cognitive and Biological Basis of Linguistic Structures*. Oxford University Press; 2013.
- Bever, TG.; Carroll, JM.; Hurtig, R. Analogy or ungrammatical sequences that are utterable and comprehensible are the origins of new grammars in language acquisition and linguistic evolution. In: Bever, TG.; Katz, JJ.; Langendoen, DT., editors. *An integrated theory of 2linguistic ability*. Crowell Press; New York: T.Y.: 1976. p. 149-182.
- Christianson K, Hollingworth A, Halliwell JF, Fernanda F. Thematic roles assigned along the garden path linger. *Cognitive Psychology*. 2001; 42:368–407. [PubMed: 11368528]
- Clark HH, Fox Tree JE. Using uh and um in spontaneous speaking. *Cognition*. 2002; 84:73–111. [PubMed: 12062148]
- Coppock, Elizabeth. Alignment in syntactic blending. *MIT Working Papers in Linguistics*. 2006; 13:239–255.
- Erickson TD, Mattson ME. From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*. 1981; 20:540–551.
- Ferreira F. The misinterpretation of noncanonical sentences. *Cognitive Psychology*. 2003; 47:164–203. [PubMed: 12948517]
- Ferreira F, Bailey KGD, Ferraro V. Good-enough representations in language comprehension. *Current Directions in Psychological Science*. 2002; 11:11–14.
- Ferreira F, Patson N. The good enough approach to language comprehension. *Language and Linguistics Compass*. 2007; 1:71–83.
- Fodor, JD.; Ferreira, F., editors. *Sentence Reanalysis*. Kluwer; Dordrecht:
- Frazier L. Two interpretive systems for natural language? In Eva Fernandez and Janet Fodor, “Grammars and Parsers: Papers from the Special Session of the CUNY Conference on Human Sentence Processing 2012”. *Journal of Psycholinguistic Research*. In press.
- Frazier, L. In: Chang, Charles; Haynie, Hannah, editors. *Processing ellipsis: A processing solution to the undergeneration problem*; *Proceedings of WCCFL*; Cascadia. 2008a;
- Frazier, L. Is “Good Enough” processing good enough?. In: Arcuri, L.; Boscolo, P.; Peresotti, F., editors. *Festschrift in honor of Ino Flores d’Arcais*. University of Padua; 2008b.
- Frazier, L.; Clifton, C, Jr.. *Construal*. MIT Press; Cambridge, MA: 1996.
- Frazier, L.; Clifton, C, Jr.. Sentence reanalysis, and visibility. In: Fodor, JD.; Ferreira, F., editors. *Sentence Reanalysis*. Kluwer; Dordrecht: 1998. p. 143-176.
- Frazier L, Clifton C Jr. Quantifiers undone: Reversing predictable speech errors in comprehension. *Language*. 2011a; 87(1):158–171.
- Frazier, L.; Clifton, C, Jr.. Dynamic interpretation: Finding an antecedent for VPE. In: Harris, Jesse A.; Grant, Margaret, editors. *University of Massachusetts Occasional Papers in Linguistics: Processing Linguistic Structure*. Vol. 38. GLSA Publishing; 2011b. p. 21-35.

- Gibson E, Bergen L, Piantadosi ST. Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *PNAS Proceedings of the National Academy of Sciences of the United States of America*. 2013; 110:8051–8056.
- Marcus, M.; Hindle, D.; Fleck, M. D-theory: Talking about talking about trees; Proceedings of the Twenty-first Annual Conference of the Association for Computational Linguistics; Cambridge, MA. 1983;
- Mehler J. Some effects of grammatical transformations on the recall of English sentences. *Journal of verbal Learning and verbal Behavior*. 1963; 2:346–351.
- Nooteboom SG, Quené H. *Journal of Memory and Language*. 2013; 69:417–428.
- Piantadosi ST, Tily H, Gibson E. The communicative function of ambiguity in language. *Cognition*. 2012; 122:280–291. [PubMed: 22192697]
- R_Development_Core_Team. A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2012.
- Sanford A, Sturt P. Depth of processing in language comprehension: Not noticing the evidence: Trends in Cognitive Sciences. 2002; 6:382–386. [PubMed: 12200180]

Figure 1.

Plot of mean naturalness ratings vs proportion undoubled interpretations, by items; double-negative items indicated (See appendix)

Figure 2.

Plot of relation between mean naturalness ratings and proportion unblended interpretations, “without his $\times$ off” sentences, Experiment 2. Open circles: Blended (doubled) items. Filled circles: Non-blended items.

Table 1

Mean proportion "acceptable", doubled items, Experiments 1A and 1B, with standard error

Experiment	Condition	Doubled	Control 1	Control 2
All Items, Negative and Non-Negative				
1A	Informal	.79 (.03)	.91 (.02)	.93 (.02)
	Formal	.71 (.04)	.91 (.02)	.93 (.02)
1B	Informal	.81 (.03)	.96 (.02)	.92 (.02)
	Formal	.72 (.04)	.93 (.02)	.83 (.03)
Non-Negative Items Only, Pooled over Informal and Formal				
1A		.83 (.02)	.88 (.02)	.94 (.02)
1B		.89 (.02)	.96 (.01)	.92 (.02)
Negative Items Only, Pooled over Informal and Formal				
1A		.48 (.06)	.99 (.01)	.89 (.04)
1B		.39 (.04)	.91 (.05)	.73 (.02)

--note: the low acceptance of Negative Control 2 items in Experiment 1B is entirely due to a very low acceptance of one item: "No social studies teachers go to see the principal."

Table 2

Mean proportion "acceptable", blend items, Experiments 1A and 1B, with standard error

Experiment	Condition	Blended	NonBlended
1A	Informal	.97 (.02)	.94 (.02)
	Formal	.97 (.02)	.97 (.02)
1B	Informal	.98 (.01)	.98 (.01)
	Formal	.88 (.03)	.96 (.02)

Table 3

Mean response times, ms, doubled items, Experiments 1A and 1B, with standard error (including all responses, both "acceptable" and "unacceptable")

Experiment	Condition	Doubled	Control 1	Control 2
1A	Informal	544 (44)	461 (36)	451 (37)
	Formal	437 (32)	434 (32)	388 (25)
1B	Informal	3204 (165)	2733 (146)	2722 (180)
	Formal	2944 (152)	2533 (121)	2569 (152)

Table 4

Mean naturalness ratings and proportions of undoubled/unblended interpretations (with standard error) for the three types of sentences, Experiment 2

Sentence type	Naturalness	Proportion undoubled/unblended
Doubled quantifier		
Doubled	3.06 (.08)	.62 (.02)
Prevent-near-culmination		
Blended	3.58 (.17)	.95 (.02)
Nonblended	3.80 (.18)	.98 (.01)
Without-his-X off		
Blended	2.28 (.10)	.34 (.03)
Nonblended	3.24 (.13)	.94 (.02)