Published in final edited form as:

J Mem Lang. 2012 July 1; 67(1): 211–223. doi:10.1016/j.jml.2012.02.011.

(Not) Hearing Optional Subjects: The Effects of Pragmatic Usage Preferences

Jennifer E. Mack^a, Charles Clifton Jr.^a, Lyn Frazier^b, and Patrick V. Taylor^a

^aUniversity of Massachusetts Amherst Department of Psychology Tobin Hall 135 Hicks Way Amherst, MA 01003

^bUniversity of Massachusetts Amherst Department of Linguistics 226 South College 150 Hicks Way Amherst, MA 01003

Introduction

A range of non-categorical constraints, which we call USAGE PREFERENCES, influence how likely a structure is to be used in a particular context. In recent years, there has been a surge of interest in usage preferences: what they are, how they work, and what they reveal about the language system. The present study contributes to a growing literature on the relationship between usage preferences and language processing. Recent research has shown that usage preferences are correlated with processing efficiency (e.g. speed and fluency) in comprehension and production (Bresnan & Ford, 2010; Brown, Savova, & Gibson, 2011; Clifton & Frazier, 2004; Gennari & MacDonald, 2009; Kaiser & Trueswell, 2004; Tily et al., 2009). In this paper, we report a different type of processing effect: when listeners are presented with acoustically degraded input, usage preferences influence what they report hearing.

Much of the existing research on usage preferences focuses on grammatical alternations, in which the grammar provides multiple potential ways to encode a message. In this paper, we investigate the alternation between expletive and null (ZERO) subjects in informal English, as illustrated in (1) (Haegeman, 1997; Mack & Fuerst, in press; Rizzi, 1994; Schmerling, 1973; Thrasher, 1977; Weir, in press; Zwicky, 2005). Other English alternations that have been central in the literature on usage preferences include the genitive alternation, shown in (2) (Hinrichs & Szendrői, 2007; Rosenbach, 2005) and the optional realization of the complementizer *that*, which appears in (2) (Ferreira & Dell, 2000; Ferreira & Hudson, 2011; Jaeger, 2010; Thompson & Mulac, 1991). Perhaps the most extensively studied alternation to date is the dative alternation (4), in which the *theme* and *goal* arguments of a three-argument verb can be realized either through a prepositional dative structure (4a) or a double-object structure (4b). Corpus studies demonstrate that the distribution of the two alternants is shaped by a wide range of usage preferences, including the information status, animacy, and length of the *theme* and *goal* arguments (Arnold, Wasow, Losongco, & Ginstrom, 2000; Bresnan, Cueni, Nikitina, & Baayen, 2007; Ruppenhofer, 2004; Thompson, 1990). Bresnan and colleagues (2007) proposed a model of the dative alternation that contained 14 independent usage preferences, including constraints related to animacy,

Address for correspondence: Jennifer E. Mack, Department of Communication Sciences and Disorders, Northwestern University, Francis Searle Building, Room 3-380, 2240 Campus Drive, Evanston, IL 60208, Jennifer-mack-0@northwestern.edu, 847-467-7591 (phone); 847-467-7377 (fax).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

length, and information structure. This model was shown to be highly successful in predicting the distribution of the two alternants in naturally-occurring data from the Switchboard corpus.

- 1)
 - a. (It) seems like people are always on their cell phones.
 - b. (It) sounds like there's probably nothing to worry about.
 - c. (It) looks like my cough isn't going away.
- 2) the president's assistants/the assistants of the president
- 3) John thinks (that) Mary is coming.
- 4)
 - a. John threw a book to Mary.
 - b. John threw Mary a book.

The dative alternation has also been the topic of recent studies investigating the relationship between usage preferences and the processing of grammatical alternants. Bresnan (2007) showed that language users can access implicit knowledge about usage preferences, at least under certain circumstances. In Bresnan's study, participants read short naturally-occurring passages and rated the relative likelihood that the passage would continue with a double-object structure or a prepositional dative. Participants' responses closely corresponded to the predictions of Bresnan et al.'s (2007) model. Other studies have tested for effects of usage preferences on processing efficiency. On the comprehension side, Bresnan and Ford (2010) showed that performance on a word-by-word lexical decision task is speeded when processing dative structures that satisfy usage preferences, relative to those that do not. Clifton and Frazier (2004) and Brown et al. (2011) focused on the effects of information status, showing that double-object structures with the preferred ordering of arguments (given *recipient* before new *theme*) are processed more quickly than dispreferred new-before-given structures. On the production side, Tily et al. (2009) reported that usage preferences affect fluency: people are more likely to produce disfluencies in dative structures that run counter to usage preferences.

These results suggest that usage preferences affect sentence comprehension and production processes. This is perhaps unsurprising, given that processing-based explanations have been proposed for some of the most common usage preferences in English and typologically similar languages, for example that animate constituents tend to appear before inanimate ones (Branigan, Pickering, & Tanaka, 2008; Gennari & MacDonald, 2009), given constituents appear before new ones (Arnold et al., 2000; Brown et al., 2011; Clifton & Frazier, 2004), and simple, short constituents appear before complex ones (Arnold et al., 2000; Hawkins, 1994; 2004). Some processing accounts involve production, centering on the notion of accessibility: constituents that are easy for speakers to access and produce, namely short, given, and/or animate constituents, tend to be produced first (Arnold et al., 2000; Bock & Warren, 1985; Branigan et al., 2008; Clifton & Frazier, 2004; Ferreira & Dell, 2000; Gennari & MacDonald, 2009). Other accounts center on comprehension, proposing that structures that are comprehended more easily – e.g., structures that minimize the length of dependencies between verbs and arguments (e.g., Gibson, 1998; Hawkins, 1994; 2004; Temperley, 2007), or that place given before new constituents (e.g., Clark & Haviland, 1977) – tend to be produced more frequently.

However, recent research suggests that the relationship between usage preferences and processing is complex. For instance, usage preferences seem not to be uniquely determined by processing factors. Bresnan and colleagues have shown that the strength of usage preferences governing the dative alternation vary across varieties of English, in the absence of any clear processing motivation (Bresnan & Ford, 2010; Bresnan & Hay, 2008). This

suggests that while usage preferences may be initially driven by processing factors, they may eventually be grammaticalized, i.e., become stored probabilistic constraints that are part of language users' knowledge. In a full circle of sorts, some researchers have argued that the production and comprehension systems have direct access to these stored probabilistic constraints, and that this accounts for effects of usage preferences on processing (e.g., Bresnan, 2007; Bresnan et al., 2007; Bresnan & Ford, 2010; Gennari & MacDonald, 2009; Jaeger, 2010).

In the present study, we investigated the question of whether usage preferences can affect what language users think they are processing. To do so, we used a novel variant of the SPEECH RESTORATION METHOD. Speech restoration occurs when listeners hear, or think they hear, segments of speech that are either absent or highly degraded in the acoustic input. Warren (1970) first noted this phenomenon, showing that listeners report hearing phonemes that have been deleted from the acoustic signal and replaced by noise. Several subsequent studies reported top-down effects on speech restoration, including effects of lexical status (more restoration in words than in non-words; Samuel, 1981; 1996) and sentential context (more restoration in predictable words than in unpredictable words; Samuel, 1981; 1996). Recently, Stoyneshka, Fodor, and Fernández (2010) used the speech restoration method to investigate the effects of prosody on syntactic ambiguity resolution in Bulgarian. In Stoyneshka and colleagues' study, participants listened to syntactically ambiguous sentences where the critical portion of a disambiguating word was replaced by noise. The prosodic contours of the sentences were hypothesized to support one parse or the other. Participants then completed a task (visual word choice, sentence repetition, or speech shadowing) that revealed which form of the disambiguating word they had heard. Stoyneshka and colleagues found that prosodic biases influenced what word participants reported hearing, indicating that prosody affected ambiguity resolution.

We are not aware of any previous studies that used the speech restoration method to study usage preferences. However, Rohde and Ettlinger (2010) recently reported effects of pragmatic expectations about causality on phoneme categorization. Like speech restoration, phoneme categorization is sensitive to lexical status (Connine & Clifton, 1987; Ganong, 1980) and sentential context (Connine, Blasko, & Hall, 1991; Isenberg, Walker, & Ryder, 1980; van Alphen & McQueen, 2001). In Rohde and Ettlinger's study, participants listened to sentences in which the main clause contained both a male and a female referent, with an acoustically ambiguous pronoun (*she/he*, abbreviated *s/he*) appearing within a causal embedded clause. Rohde and Ettlinger showed that pragmatic expectations influenced which pronoun participants reported hearing. When the context was biased towards co-reference with the matrix subject, participants often reported hearing pronouns consistent with that interpretation (e.g., *she* in 5a). Likewise, they tended to report hearing pronouns that co-referred with the matrix object when context supported this interpretation (e.g., *he* in 5b).

- 5)
 - a. Abigail annoyed Bruce because s!he was in a bad mood.
 - b. Joyce helped Steve because s!he was working on the same project.

These findings suggest that non-categorical pragmatic constraints can influence what people report hearing when they are presented with acoustically ambiguous input. In the present study, we test whether pragmatic usage preferences affect which of two grammatical alternants is heard. We focus on the expletive/zero subject alternation, which we will describe and analyze below.

The Expletive/Zero Subject Alternation

The expletive/zero subject alternation involves the optional realization of an expletive (non-referential, or 'dummy') subject. The alternation affects the full range of expletive subjects

in English: extraposition *it*, which appears when a sentential argument is extraposed to post-verbal position (6a–c), presentational/existential *there* (6d–e), appearing in sentences that introduce a new referent or comment on the (non)existence of an entity, and weather *it*, which appears with weather verbs (6f).

- 6)
 - a. (It) looks like everyone is managing to keep up.
 - b. (It) turns out you can't do that in Texas. (Schmerling, 1973: 582)
 - c. (It) won't be too difficult to reconstruct his argument. (Thrasher, 1977: 44)
 - d. (There) isn't much we can do about it. (Thrasher, 1977: 44)
 - e. (There) must be an accident up ahead. (Schmerling, 1973: 582)
 - f. (It) always rains on Mondays. (Weir, in press: his 4f)

The expletive/zero alternation itself has been the focus of little previous work (although see Mack & Fuerst, in press; Zwicky, 2005), but it is part of a broader phenomenon that has been studied extensively. In informal spoken and written English, referential subjects can also be omitted (7a–b). In addition, particularly in spoken English a wide variety of sentence elements can be omitted sentence-initially, including auxiliaries (7c), determiners (7d), and even word-initial syllables (7e), a phenomenon that has been called INITIAL DELETION.

- 7)
 - a. (I) won't be in the office tomorrow. (Weir, in press: his 4a)
 - b. A: Am I invited to the party? B: (You) must be, surely. (Weir, in press: his 4c)
 - c. (Have) you seen Tom? (Napoli, 1982: 85)
 - d. (The) paper boy's here. (Napoli, 1982: 85)
 - e. (Has the Pro-) 'fessor arrived yet? (Napoli, 1982: 85)

These phenomena have been analyzed at the prosodic, syntactic, and pragmatic levels. Null subjects in spoken English have been argued to result from a prosodic process: the deletion of unstressed material from the left edge of a prosodic phrase (Gerken, 1991; Napoli, 1982; Thrasher, 1977; Weir, in press). Proponents of the prosodic analysis have supported it with three major observations. First, as illustrated above, initial deletion can apply to many types of linguistic elements, some of which (e.g., 7e) are word-internal, and thus prosodic, but not syntactic, constituents. Second, subject deletion in spoken English can only occur sentence-initially, as illustrated by the contrast between (8a) and (8b). Finally, unstressed constituents such as auxiliaries cannot stand alone after subject deletion (9a); they must either be deleted along with the subject (9b) or be cliticized to a stressed word (9c).

- 8)
 - a. (It) seems like Tom is happier than usual today.
 - b. Today *(it) seems like Tom is happier than usual.
- 9)
 - a. *(It) is pretty neat that she made the cake herself.
 - b. (It's) pretty neat that she made the cake herself.
 - c. (It)'s pretty neat that she made the cake herself.

These observations suggest that subject omission in spoken English, including the expletive/zero alternation, is a prosodic process. Weir (in press) provides an Optimality-Theoretic analysis of initial deletion which hinges on the constraint STRONGSTART, which penalizes structures in which the leftmost constituent is less prosodically prominent than the one that

follows. In certain circumstances, this licenses omission of unstressed material at the left edge of an intonational phrase.

In contrast, Haegeman and colleagues have proposed syntactic analyses of subject omission in written English, often called ‘diary drop’ (Haegeman 1990; 1997; 2007; Haegeman & Ihsane, 1999, 2001); while these analyses have evolved over time, they share the assumption that the distribution of null subjects is defined in terms of syntactic rather than prosodic structure. Haegeman supports this claim with attested examples of null subjects that are not sentence-initial, e.g. (10) below. In addition, as Weir (in press) points out, deletions at the syllabic level, e.g. (7e) above, are dispreferred in written language. These data demonstrate that the distribution of null subjects differs in written and spoken English, supporting a prosodic account of subject omission in spoken language but potentially favoring a syntactic account in writing. In the present paper, we will focus on the expletive/zero alternation in spoken language, and will thus assume that the alternation is conditioned by prosodic constraints.

- 10)** So $\emptyset$ shall now stop writing for the day. (Haegeman, 1990: 164, citing the diary of Virginia Woolf)

In addition, it has often been suggested that the expletive/zero alternation is subject to pragmatic constraints. Most accounts of subject omission note that it primarily occurs in informal speech and writing. In Weir’s (in press) analysis, the effects of register are captured formally through different constraint rankings in formal and informal registers, with the result that subject omission is licensed only in informal contexts.

Furthermore, some previous researchers have claimed that the expletive/zero alternation is conditioned by subtle pragmatic constraints above and beyond the effects of register (Mack & Fuerst, in press; Schmerling, 1973; Thrasher, 1977; Zwicky, 2005). For instance, it has been claimed that zero sentences are more “subjective” than sentences with expletive subjects (Mack & Fuerst, in press; Thrasher, 1977; Zwicky, 2005). Zwicky (2005) observes a subtle pragmatic distinction between (11a) and (11b), with the latter being more closely tied to the perspective of the speaker. The subjectivity constraint manifests itself more obviously in sentences with overt experiencers, e.g. (12). While expletive-subject sentences permit both first-person and third-person experiencers, zero sentences are fully natural only with first-person experiencers (Mack & Fuerst, in press; Schmerling, 1973).

- 11)** **a.** It’s odd that Mary showed up.
 b. Odd that Mary never showed up.
- 12)** **a.** It looks to me/Sam like things are going pretty well.
 b. Looks to me/?Sam like things are going pretty well.

We propose that subjectivity is one facet of a more general pragmatic constraint on the expletive/zero alternation, namely that zero sentences typically express IMMEDIATE JUDGMENTS. An immediate judgment is a judgment that the speaker formed just before the sentence was uttered. This idea is in line with Schmerling’s (1973: 582–583) observation that zero sentences involve “some elusive element of spontaneity and impulsiveness.” The immediate judgments hypothesis accounts for the observation that zero sentences tend to be subjective: we can express our own opinions, but not those of others, immediately after they are formed. We call this aspect of immediacy PERSONAL IMMEDIACY. The hypothesis also predicts that speakers should use zero sentences relatively frequently when expressing recently-formed judgments, and less frequently to express judgments formed further in the past. We refer to this aspect of immediacy as TEMPORAL IMMEDIACY. The immediate

judgments hypothesis was tested with a corpus study (Study 1) and a speech restoration study (Study 2).

Study 1

We tested the temporal immediacy hypothesis using the spoken portion of the Corpus of Contemporary American English (COCA; Davies, 2008–). We were unable to test the personal immediacy hypothesis because our initial searches failed to find any sentences with third-person experiencers.¹

Materials and Methods

When the searches were performed (September 7, 2011), the spoken portion of the COCA consisted of approximately 90 million words from TV and radio programs aired over the past two decades. We searched for expletive-subject and zero sentences headed by an evidential verb (*seem, look, sound*) that selected for a sentential complement headed by the particle *like* or *as*. Specifically, we searched for sentence-initial strings of the form ‘(it) looks|seems|sounds like|as’ and ‘(it) looked|seemed|sounded like|as’, where the symbol | indicates alternatives. After performing the searches, we inspected each result, excluding repeated tokens, sentences that did not have the correct structure (e.g., sentences with nominal complements), and sentences with overt experiencers in the matrix clause. There were 989 remaining tokens that were entered into further analyses.²

While the expletive/zero alternation occurs in a far wider range of environments (see example 6), we focused on this particular environment because it corresponds closely to our stimuli in Study 2. Examples of expletive (13) and zero (14) sentences returned by the corpus searches appear below.

- 13)
 - a. It looks like March Madness will ensue with only one undefeated team.
 - b. It sounds like you’ve made up your mind.
 - c. It seemed like everyone had a tear in their eye except him.
 - d. It looked as if the pilots deliberately targeted these government buildings.
- 14)
 - a. Sounds as if he had been involved with another woman before divorcing his wife.
 - b. Looks like Philadelphia is in the lead.
 - c. Seemed like it was too good to be true.
 - d. Looked like you were going to be there for a while.

Data Analysis

To test the predictions of the immediate judgments hypothesis regarding temporal immediacy, we calculated the frequency with which expletive and zero subjects co-occurred with present-tense and past-tense main verbs. Additionally, we calculated the distribution of expletive and zero responses for each of the three verbs included, as verb-specific usage preferences have previously been found in a range of grammatical alternations, including the

¹These searches were for sentence-initial strings of the form ‘(it) looks|looked|seems|seemed|sounds|sounded to me|us like|as’ and ‘(it) looks|looked|seems|seemed|sounds|sounded to him|her|them like|as’, with the symbol | indicating alternatives.

²Note that our sample most likely contains some referential subjects, both overt (*it*) and null. However, it is often very difficult to determine whether an ambiguous subject is referential in context, so we did not attempt to do so. Thus, our results may characterize subject omission generally, which subsumes the expletive/zero alternation.

dative alternation (Bresnan et al., 2007; Bresnan & Ford, 2010) and complementizer omission (Jaeger, 2010).

We used logistic regression to test for effects of temporal immediacy and verb, using the *glm* function in R (R Development Core Team, 2011). Both predictor variables were contrast (i.e., sum) coded, so that their effects can be interpreted as main effects, marginalized over the other factors. We compared four models of the data: Model 1 contained only a main effect of tense, Model 2 a main effect of verb, Model 3 main effects of tense and verb, and Model 4 main effects of tense and verb and a tense x verb interaction. A more complex model was selected only if it was a better fit to the data than a simpler model, using the χ^2 test with a threshold of $p < .15$.³ Significance statistics for the selected model were provided by the *glm* function.

Results

The distribution of expletive and zero sentences with present- and past-tense verbs is summarized in Table 1. In our data, expletive subjects occurred more often than zero subjects and present-tense verbs were more frequent than past-tense verbs. Critically, verb tense conditioned the alternation: with present-tense verbs, the ratio of expletive to zero subjects was 4.3 to 1, while with past-tense verbs the ratio increased to 8.6 to 1. This finding supports the temporal immediacy hypothesis. The corpus results also support the hypothesis that the immediacy constraint is a usage preference, rather than a categorical constraint: while zero sentences are rarely used to express non-immediate judgments, it does occasionally occur (e.g., 14 c–d).

Table 2 summarizes the frequency of expletive and zero sentences with each of the three main verbs included in our corpus searches. The results suggest that the main verb does condition the alternation, with *seem* having the strongest bias towards expletive subjects (with an expletive: zero ratio of 9.1:1), followed by *look* (4.3:1), then by *sound* (2.4:1).

The model comparison procedure led to the selection of Model 3, containing main effects of tense and verb, but no interaction between these factors. The parameters of this model are summarized in Table 3. Positive coefficient estimates indicate that the named level of a factor increases the probability of an expletive subject, relative to the mean across all levels of the factor; negative coefficient estimates indicate a greater likelihood of zero sentences. Thus, the results show that zero sentences are significantly more frequent in present-tense contexts than they are in the data generally ($p < .05$). The results for verb indicate that *seem* is associated with a higher rate of expletive subjects ($p < .0001$) and *sound* with a higher rate of zero subjects ($p < .0001$), relative to the mean across verbs.⁴

Discussion

The corpus results indicate that the expletive/zero alternation is conditioned both by temporal immediacy and by the identity of the main verb. Consistent with the temporal immediacy hypothesis, zero sentences were found to occur more frequently in temporally immediate contexts than in non-immediate contexts. In addition, we also observed verb-specific effects on the alternation, with *seem* occurring more frequently with expletive subjects and *sound* more frequently than zero subjects than the mean across verbs. This finding is consistent with previous corpus studies that report verb-specific effects on grammatical alternations (e.g., Bresnan et al., 2007; Bresnan & Ford, 2010; Jaeger, 2010).

³This liberal threshold was employed because an additional significance test was performed on the selected model.

⁴We report contrasts for *seem* and *sound*, but not *look*, because in a three-level comparison with contrast coding, there are only two degrees of freedom, and thus only two independent contrasts. We chose *seem* and *sound* for these comparisons because they have the highest and lowest rates of expletive subjects, respectively.

Study 2: Speech Restoration

With corpus data providing support for the immediate judgments hypothesis, we aimed to investigate whether immediacy affects the processing of the expletive/zero alternation. We first conducted a pilot study in which participants rated the acceptability of auditorily-presented past- vs. present-tense sentences with expletive vs. zero subjects. Sentences with zero subjects received overall lower acceptability ratings than ones with expletive subjects (3.83 vs. 3.99 on a five-point scale), but the difference was not affected by the tense (temporal immediacy) of the sentence. Although the fact that the mean ratings were close to the end of the scale means that we cannot conclude that participants' conscious evaluations of zero sentences are insensitive to immediacy, the pilot data did not encourage use of explicit ratings. Instead, we turned to an implicit task. In Study 2, we used speech restoration to test the hypothesis that temporal and personal immediacy would influence whether participants report hearing an expletive or zero subject.

Materials and Methods

Participants—48 undergraduates from the University of Massachusetts-Amherst participated in the experiment in exchange for course credit. All were native speakers of English with normal hearing. The study was approved by the Institutional Review Board at the University of Massachusetts-Amherst.

Stimuli—We constructed 24 three-sentence dialogues, each concluding with a sentence permitting an optional expletive subject (see Appendix). Three main verbs were used (*seem*, *sound*, *look*), each of which appeared in 8 dialogues. There were four versions of each dialogue, crossing two factors, temporal immediacy and personal immediacy. We manipulated temporal immediacy through verb tense and sometimes temporal modifiers: immediate contexts included present-tense verbs and modifiers (*present-tense* condition; 15a), whereas non-immediate contexts contained past-tense verbs and modifiers (*past-tense* condition; 15b). Personal immediacy was manipulated through an adjunct prepositional phrase, which either identified the speaker (*first-person* condition) or a third party (*third-person* condition) as the experiencer.

- 15)**
- a.**
- A** Joe and I are studying like crazy for the German exam.
- B** Oh yeah?
- A** It seems to me/him like there are thousands of words to learn.
- b.**
- A** Joe and I were studying like crazy for the German exam.
- B** Oh yeah?
- A** It seemed to me/him like there were thousands of words to learn.

The dialogues were recorded in a soundproof room by two native speakers of English, one male and one female. The male speaker always produced the first and last lines of each dialogue, including the crucial final sentence. We spliced the dialogues so that linguistically identical sentences were acoustically identical across conditions. With a few exceptions meant to improve clarity and prosodic consistency across items, we did not splice linguistic material within sentences.

We then distorted the expletive subject of the final sentence of each dialogue by applying a distorting filter that we created in Adobe Audition, combined with an overlay of white noise.

⁵ The result, to our ears, was acoustically ambiguous between a “noisy” expletive *it* and pure noise, containing no speech. In order to ensure that the ambiguous segments were comparable across conditions, we measured the length of each token of *it* before applying the distorting filter. As shown by Table 4, the distorted segments were of approximately equal length across conditions ($F(3) = 1.07, p = .37$).

To make the placement of “noise” less predictable to participants, we also distorted a second segment of each experimental dialogue. Half of the time the distorted segment was part of the background noise that preceded or followed an utterance, and half of the time the distorted segment was a syllable of speech that was not crucial to the experimental manipulation (i.e., we never distorted syllables that contained markers of personal or temporal immediacy). The position of the second distortion was always identical within items.

The experimental dialogues were interspersed with 60 filler dialogues, each two or three sentences long and concluding with an utterance produced by the male speaker. The filler dialogues each contained one, two, or three distorted segments; half of the distortions were applied to syllables of speech, and half to short segments of background noise. The experimental and filler dialogues were presented in a randomized order, with scripts counterbalanced so that every subject saw one version of each experimental dialogue, distributed evenly across the four conditions.

Procedure

Participants sat in a small soundproof room in front of a computer monitor. Auditory stimuli were presented through two speakers, one each in the left and right sides of the room. We used a modified version of Linger (Rohde, 2003) to present the stimuli and record subjects’ verbal and written responses.

Participants were told that they would be listening to short dialogues that would occasionally be noisy. They were instructed to listen to each dialogue carefully and then, following the appearance of a visual prompt, repeat the final sentence of the dialogue. They were asked to repeat exactly what they had heard, or if they were not sure, what they thought the speaker had probably said. After repeating the sentence, participants were also asked to type the first five or six words of the same sentence. When they indicated that they were finished typing, the next dialogue began. Before beginning the experiment, participants completed a 7-item practice session that familiarized them with the task. To prevent participants from responding prescriptively (and thus possibly avoiding zero sentences), we emphasized that the participants should repeat what they had heard verbatim, including informal sentences and sentences containing “mistakes,” such as disfluencies. For the same reason, we included several non-standard and/or disfluent utterances in the practice and filler dialogues.

We chose a sentence repetition task for two reasons: it had previously been shown to be effective at capturing higher-level effects on restoration (Stoyneska et al., 2010) and it kept the linguistic contrasts that we were manipulating relatively obscure. The original purpose of the written task was to provide us with analyzable responses in the case that phonetic reduction in the verbal responses made it difficult to determine whether the participant had

⁵The filter reduced the amplitude of high-amplitude portions of the input signal and increased the amplitude of low-amplitude portions of the signal. Specifically, the filter was a transformation with the following coordinates, connected linearly, where the first member of each pair is the input amplitude and the second member the output amplitude in db: [−60, −60; −54, −15; −41, −15; 0, −45]. The overlay of white noise had an intensity value of 3.9 on the scale used by Adobe Audition (which appears as approximately −19 db average RMS power; this can be compared to −20 db average RMS power for the full length of a typical sentence).

produced a subject. However, as this proved not to be a problem, we will focus on the spoken responses in this paper.

Data Analysis

We excluded responses that contained errors relevant to the experimental manipulation: omission of the matrix clause, disfluency at the beginning of the utterance, or errors in verb tense or the person of the experiencer, the markers of immediacy. In the spoken responses, all 48 participants showed strong performance on the task, repeating the final sentence of the dialogue accurately (per the definition provided above) on at least 75% of experimental trials. 5.1% of the spoken responses were excluded due to error. Due to a greater number of errors in the written responses, we restricted the analysis to 44 participants who showed strong overall performance on the task (75% on experimental items). For the included participants, 5.9% of responses were excluded due to error. The remaining responses were coded for the presence or absence of an expletive subject. For the spoken responses, the coding was done by two researchers working independently, one of whom was unaware of the experimental predictions. There was 98% agreement between the two researchers. When disagreement arose, we used spectrograms generated in Praat (Boersma, 2001) to code the utterance. Utterances with initial vowel formants were coded as expletive subjects; those that lacked vowel formants were coded as zero subjects.

We tested for effects of temporal immediacy (tense), personal immediacy (person), and verb on the rate of expletive-subject responses with mixed-effects logistic regression, using the *lmer* function from the languageR package (Baayen, 2011). All predictor variables were contrast coded. We compared seven models with increasingly complex fixed effects structures: Model 1 had a main effect of tense, Model 2 had main effects of tense and person, Model 3 had main effects of tense and person and a tense x person interaction, Model 4 had main effects of tense, person, and verb Model 5 had main effects of all three predictors and a person x verb interaction, Model 6 had main effects of all three predictors and a tense x verb interaction, and Model 7 had main effects of all three predictors and a three-way interaction.⁶ All models tested had maximal random effects structures, i.e., participant- and item-specific random slopes and intercepts for all fixed effects contained in the model. A more complex model was chosen over a simpler model only if it improved the fit to the data, using the χ^2 test with a threshold of $p < .15$. To determine whether a given fixed effect improved model fit, we compared a model containing that fixed effect to an otherwise identical model that omitted it. The *lmer* function provided significance statistics for the coefficients of the fixed effects of the selected model.

Results

Spoken Responses—Overall, participants' spoken responses included an expletive subject 68.3% of the time. As Table 5 illustrates, immediacy influenced the likelihood that the participants would report hearing an expletive subject. Participants were more likely to restore the expletive subject in temporally non-immediate (i.e., past-tense) contexts than in immediate present-tense contexts. They were also more likely to report hearing expletive *it* in personally non-immediate contexts (third person experiencers) than in personally immediate contexts (first person experiencers).

The model comparison procedure led to the selection of Model 4, which contained fixed effects of tense, person, and verb, but no interactions, with participant- and item-specific random slopes and intercepts for each predictor. Table 6 summarizes the fixed effects of this model. As in Study 1, positive coefficient values indicate an increased likelihood of

⁶Models 6 and 7 failed to converge, and thus were not considered further.

producing an expletive subject, while negative coefficient values reflect an increased rate of zero sentences. Thus, the model indicates that present-tense sentences elicited significantly more zero responses than past-tense sentences ($p < .05$). There was also a marginally significant effect of person, with first-person experiencers eliciting more zero responses than third-person experiencers ($p = .059$). In addition, there was a significant effect of verb: participants restored *it* significantly more frequently in sentences with *seem* ($p < .01$) and marginally less frequently in sentences with *sound* ($p = .081$), relative to the mean across verbs.

While the effect of tense was fully significant in the selected model, the effect of person only approached significance. We hypothesized that the effect of person may have been attenuated due to conflict between cues to immediacy in the two “mixed” conditions (present-tense verb with third-person experiencer and past-tense verb with first-person experiencer). For this reason, we performed an additional analysis using only the two conditions that provided consistent, “extreme” cues to immediacy (present-tense verb with first-person experiencer, past-tense verb with third-person experiencer). We tested four models of this subset of the data, all with maximal random effects structures. Model 1 included a main effect of immediacy, Model 2 a main effect of verb, Model 3 main effects of immediacy and verb, and Model 4 main effects of immediacy and verb and an immediacy $\times$ verb interaction. Model selection procedures were the same as that used for the full data set. The parameters of the selected model are summarized in Table 7. The model includes significant main effects of immediacy (more zero responses in immediate than in non-immediate contexts; $p < .001$) and verb (more *it* responses for *seem* ($p < .01$) and more zero responses for *sound* ($p < .001$) than the mean across verbs), but no interaction. This suggests that immediacy significantly conditions the expletive/zero alternation when consistent cues are provided.

Written Responses—Interestingly, the overall rate of restoration was higher in our participants’ written responses (75.6%) than in their spoken responses (68.3%). This may be due to effects of register or prescriptive norms, with participants responding more prescriptively in writing than in speech. Alternatively, as the written responses were produced after the spoken responses, it is possible that participants’ memories of the sentences to be repeated were less accurate, resulting in a greater bias to respond with an expletive subject, the more frequent of the two alternants.

The distribution of written responses (Table 8) was similar to that of the spoken responses. There was a numerically higher rate of restoration of *it* in non-immediate contexts, both temporal (past-tense verb) and personal (third-person experiencer). However, the effects of immediacy were not significant in models with maximal random effects structures (random slopes and intercepts for participants and items).

The attenuated effects of immediacy in the written responses may be due to reduced power (a slightly lower number of participants and included responses) and greater variability in responses across participants and items. One possible cause of increased variability is that tense and person may be less reliable markers of immediacy in written language than in speech. In fiction, for instance, it is common to use past-tense narration to describe events as they unfold. In addition, stories are often presented from the perspective of a third-person narrator whereas in speech, shifting perspective away from the speaker is difficult (though not impossible; see e.g., Harris & Potts, 2009). For this reason, it may be possible in written language to interpret sentences with past-tense verbs and/or third-person experiencers as immediate. The results of a small corpus study on experiencers are consistent with this possibility. We searched the spoken (90 million words) and fiction (85 million words) sections of the COCA for expletive and zero sentences with overt pronominal experiencers:

first person (*me, us*) or third person (*him, her, them*). We found that third-person experiencers were more common in writing than in speech: 19% (8/42) of the written sentences had third-person experiencers, whereas 0% (0/60) of spoken sentences did. This difference is significant despite the small sample size (Fisher's Exact Test, $p < .001$). This is consistent with the hypothesis that third-person experiencers are sometimes immediate in writing but not in speech. Inconsistency in how immediacy is marked across genres may have led to greater variability in participants' written responses.

Discussion

The main finding of Study 2 was that when presented with an acoustically ambiguous sentence, listeners tended to report hearing zero sentences more often in immediate contexts. We also found a significant effect of verb, with *seem* eliciting relatively more restoration of *it* and *sound* marginally more zero responses than the mean across verbs. Both findings are consistent with the corpus results from Study 1. Thus, the results of the restoration study lend further support to the hypothesis that immediacy shapes the expletive/zero alternation. More generally, these results demonstrate that usage preferences can influence what structure listeners ultimately assign to ambiguous input.

General Discussion

Immediacy and the Expletive/Zero Alternation

In this paper, we have presented corpus and experimental evidence in support of the hypothesis that the expletive/zero alternation is shaped by immediacy, with zero sentences typically expressing immediate judgments. Study 1, a corpus analysis, demonstrated that zero sentences are disproportionately rare with past-tense main verbs, consistent with the hypothesis that they are largely constrained to present-tense, temporally immediate environments. Study 2, a speech restoration experiment, revealed effects of immediacy that were in line with the predictions of the immediate judgments hypothesis, as well as the corpus data: participants were more likely to report hearing an expletive subject in non-immediate contexts, both temporal and personal, than in immediate contexts. We thus conclude that the expletive/zero alternation is indeed shaped by immediacy.

This raises the question of whether the immediacy constraint on zero sentences may emerge from a more general constraint on language production or comprehension. One possibility is that the immediacy constraint stems from a production-based tendency for speakers to reduce constituents that are relatively easy to formulate. Immediate judgments, both personal and temporal, may be easier to formulate than non-immediate judgments because they express thoughts that the speaker has just formed, and that are thus highly accessible. For this reason, they may undergo reduction (i.e., subject omission) more frequently than non-immediate judgments. Previous studies have suggested that phonetic and syntactic reduction occurs more frequently before accessible entities than before less accessible entities. For example, Fox Tree and Clark (1997) show that speakers tend to phonetically reduce *the* to *th[ə]* before an accessible noun, while using the unreduced form *the[ij]* when they are experiencing difficulties accessing or producing the following noun. At the syntactic level, some researchers have previously observed effects of accessibility on complementizer omission, with *that* being omitted more frequently when the subject of the complement clause is accessible (Ferreira & Dell, 2000; Jaeger et al., 2010). There has been relatively little work on the effects of the accessibility of entire messages on syntactic reduction. However, Ferreira and Hudson's (2011) study on complementizer omission suggests that messages that are easy to formulate may be prone to reduction. In a series of experiments involving interactive dialogues, participants produced sentences in which one interlocutor attributed emotions to another (e.g. *I/You/She believe(s) (that) I/you/she am/are/*

is angry). The authors found that the rate of complementizer omission was affected both by the identity of the matrix subject (more omission with first-person than with second- and third-person subjects) and the embedded subject (more omission with second- and third-person subjects than with first-person subjects). Ferreira and Hudson attribute these effects to the relative difficulty of formulating different types of messages about emotional attribution. They argue that it is easier to adopt one's own perspective than that of someone else, leading to increased rates of complementizer omission in sentences with first-person matrix subjects, and that it is easier to attribute emotions to others than to oneself, resulting in more complementizer omission with second- and third-person embedded subjects. The matrix-subject effect reported by Ferreira and Hudson may be related to the personal immediacy effect that we found for zero sentences.

Alternatively, it may be possible to account for immediacy effects using the notion of predictability. Some researchers have proposed predictability-based accounts of reduction, both phonetic (e.g., Aylett & Turk, 2004) and syntactic (Jaeger, 2010), in which reduction targets constituents that are predictable given the prior linguistic context. For example, Jaeger (2010) shows that predictability affects the likelihood of complementizer omission: *that* is omitted more frequently following verbs that often take sentential complements than following verbs that do so less frequently. Jaeger argues that this reflects a preference to distribute information evenly across the linguistic signal, which could result either from constraints on comprehension or production. Could predictability be the source of the effects of immediacy on the expletive/zero alternation? Possibly, but in order to evaluate this hypothesis we must first determine how to quantify the predictability of expletive *it* (or the extraposition structures in which it occurs). This is difficult given that it appears sentence-initially. One possibility is to define predictability at the pragmatic level. For example, discourses about the "here and now" (i.e., immediate contexts) may raise the likelihood that the speaker will utter a subjective judgment, which often take the form of an extraposition structure headed by an epistemic verb. If extraposition structures are indeed relatively predictable in immediate contexts, it could account for the observation that they often undergo reduction (i.e., subject omission) in these contexts. More research is needed to determine whether this analysis is a viable alternative to the accessibility account.

Using Speech Restoration to Study Grammatical Alternations

Previously, Stoyneshka et al. (2010) used speech restoration in order to study effects of prosody on syntactic ambiguity resolution. The results of the present study demonstrate that speech restoration may also be a useful tool for studying the effects of usage preferences on grammatical alternations. In fact, this methodology may be particularly well-suited to the investigation of usage preferences because it avoids the potential problem of usage preferences weakening under exposure to balanced stimulus sets (for discussion of this issue, see Jaeger, 2010). The speech restoration method addresses this problem by presenting participants with acoustically ambiguous stimuli and allowing them to generate the distributions.

The present study is one of several studies that have found effects of sentence- or discourse-level constraints on the processing of speech, such as speech restoration or phoneme categorization (Connine, 1987; Connine et al., 1991; Isenberg et al., 1980; Rohde & Ettlinger, 2010; Samuel, 1981; 1996; Stoyneshka et al., 2010; van Alphen & McQueen, 2001). Of central interest in many of these studies is whether such effects are perceptual or post-perceptual in nature. Samuel (1981; 1987; 1996) developed a discrimination task designed to distinguish between these two possibilities. In these studies, listeners heard words or sentences containing either a segment that was replaced by noise or a segment to which noise was simply added, and were asked to judge whether the segment was intact. Samuel was thus able to disentangle listeners' ability to discriminate noise-added from

noise-replaced sounds, a perceptual measure, from any post-perceptual bias to report stimuli as intact in particular contexts. Samuel (1981; 1996) reported lexical effects on both perceptual and post-perceptual measures, but sentential context was only associated with post-perceptual effects. Similarly, Connine (1987) and van Alphen and McQueen (2001) provide evidence that sentence-level effects on phoneme categorization are also likely to be post-perceptual.

The present speech restoration study was not designed to test these questions. However, it is clear that at least some of the observed effects of immediacy on restoration were post-perceptual. In the experiment, the sole marker of personal immediacy, the person of the experiencer, appeared two words after the subject (e.g., *(It) seems to me ...*), suggesting that the effects of personal immediacy cannot have been perceptual. As there were markers of temporal immediacy both before and after the subject, we cannot draw conclusions about the processing stage affected by this contrast. However, given Samuel's (1981; 1996) findings, it seems likely that effects of usage preferences on speech restoration are post-perceptual. They are likely involved in the process of generating a final parse of an initially ambiguous structure. This hypothesis could be tested in future research using Samuel's discrimination paradigm: if effects of usage preferences on restoration are indeed post-perceptual, we would not expect them to influence discrimination between noise-added and noise-replaced segments.

Usage Preferences and Language Comprehension

The role of usage preferences in the processing system is not yet well understood. While many usage preferences are thought to be motivated by processing constraints, either synchronically or diachronically, it has been claimed that some usage biases do not have discernible processing motivations (e.g., Bresnan & Ford, 2010; Bresnan & Hay, 2008). Thus, it is not yet clear how usage constraints affect the processing system, particularly in the domain of comprehension. Many questions remain, among them: Are effects of usage preferences on comprehension due to stored probabilistic constraints governing the use of grammatical structures, or to more general constraints governing the comprehension or production systems? What stages of comprehension (i.e., prediction of upcoming structure, revision of previous interpretations) are affected by usage preferences?

Previous research has shown that usage preferences can affect processing speed in comprehension, with structures that satisfy usage preferences being processed faster than those that do not (Bresnan & Ford, 2010; Brown et al., 2011; Clifton & Frazier, 2004; Gennari & MacDonald, 2009; Kaiser & Trueswell, 2004). However, it is still an open question whether these effects are due to stored probabilistic information, constraints on comprehension, and/or sensitivity to constraints on production. Some have argued that the grammar contains stored information about the probabilities that particular words and structures will appear in a given context (Bresnan, 2007; Bresnan & Ford, 2010; Bresnan & Hay, 2008; Manning, 2003; Wasow, 2009). If usage preferences are built into the grammar, then they are naturally part of the comprehension system. Alternatively, preferred structures may be processed more quickly due to constraints on comprehension, such as greater ease processing predictable structures (e.g., Jaeger, 2010; Levy, 2008b) or structures with shorter dependencies (e.g., Gibson, 1998; Hawkins, 1994, 2004; Temperley, 2007). A third possibility is that the comprehension system is sensitive to constraints on production. Consistent with this idea, Frazier and Clifton (2011) provided evidence that listeners often repair speech errors that are motivated by constraints on production. There is an ongoing debate about whether production influences comprehension directly or whether it is mediated through exposure to the output of the production system (see discussion in Gennari & MacDonald, 2009; Hemforth et al., 2010). Direct effects of production may arise from

activation of the production system during comprehension (Pickering & Garrod, 2007) or shared processing resources between production and comprehension (Chang, Dell, & Bock, 2006).

The present results are potentially compatible with all three possibilities. Proponents of probabilistic grammars might explain the results of Study 2 by arguing that comprehenders have access to co-occurrence probabilities between subject omission and markers of immediacy (i.e., tense and temporal markers, the person of the experiencer). Alternatively, effects of immediacy on restoration may be due to general constraints on comprehension (e.g., a preference for reduction of predictable material) or production (e.g., a tendency to reduce accessible constituents).

We now turn to the issue of how usage preferences affect different aspects of comprehension. As mentioned earlier, there is evidence that contextually-preferred structures are processed faster than dispreferred structures (Bresnan & Ford, 2010; Brown et al., 2011; Clifton & Frazier, 2004; Gennari & MacDonald, 2009; Kaiser & Trueswell, 2004). The results of Study 2 demonstrate that usage preferences can also influence which syntactic structure participants report hearing when presented with noisy input. This suggests that the comprehension system takes non-categorical constraints into account when resolving uncertain input. Previously, Levy (2008a, 2011) proposed a computational model that is capable of such effects. Levy's model was initially developed in order to account for local coherence effects, in which comprehenders generate syntactic structures that are compatible with local but not global syntactic constraints. For instance, in sentences such as (16a) they sometimes interpret *tossed* as a finite verb rather than a past participle (Tabor, Galantucci, & Richardson, 2004). Levy (2008a) proposed that such effects arise due to uncertainty about the input: for example, a comprehender may question whether *at* was really *and* or *as*, which would license the verb interpretation. Generating this interpretation in turn strengthens the comprehender's uncertainty about the prior input. In a recent eyetracking study, Levy, Bicknell, Slattery, and Rayner (2009) provided experimental support for this claim. In that study, participants read sentences such as (16a), in which the preposition had orthographic near-neighbors that would license the verb interpretation, and sentences such as (16b), in which the preposition had no near-neighbors of this sort. The results revealed evidence of greater processing difficulty at the critical word (*tossed*) in (16a), suggesting that comprehenders were revising an initially uncertain parse.

- 16)** **a.** The coach smiled at the player tossed the Frisbee.
 b. The coach smiled toward the player tossed the Frisbee.

A key aspect of Levy's model is that it allows the resolution of uncertain input on the basis of subsequent input. This feature is particularly relevant to our speech restoration results. As discussed above, the cues to personal immediacy in our stimuli always followed the ambiguous subject. Therefore, the effects of personal immediacy on restoration must have been postdictive, with the comprehender revising an initially ambiguous or uncertain parse of the input. The results thus suggest that usage preferences play a role in syntactic revision processes, in addition to their role in the prediction of upcoming syntactic structures.

To conclude, the results of the present study suggest that usage preferences affect what grammatical structure listeners assign to acoustically ambiguous input. Resolving ambiguous input is a significant aspect of language processing: while the noisy dialogues presented to our participants were artificial, real conversations often take place under noisy conditions as well. Therefore, further research on the effects of non-categorical constraints on grammatical ambiguity resolution may lead to a better understanding of the role these constraints play in the language processing system.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Desiree Svegliati for assistance with Study 2. We are also grateful for helpful comments from the UMass Psycholinguistics Group and the audience at CUNY 2011. This research was supported by an ARRA supplement to grant HD18708 to Charles Clifton and Lyn Frazier.

References

- Arnold JE, Wasow T, Losongco A, Ginstrom R. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*. 2000; 76:28–55.
- Aylett MP, Turk A. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*. 2004; 47:31–56. [PubMed: 15298329]
- Baayen, RH. *languageR: Data sets and functions with “Analyzing linguistic data: A practical introduction to statistics”*. 2011. <http://CRAN.R-project.org/package=languageR>
- Bock JK, Warren RW. Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*. 1985; 21:47–67. [PubMed: 4075761]
- Boersma P. Praat, a system for doing phonetics by computer. *Glott International*. 2001; 5:341–345.
- Branigan HP, Pickering MJ, Tanaka H. Contributions of animacy to grammatical function assignment and word order during production. *Lingua*. 2008; 118:172–189.
- Bresnan, J. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In: Featherston, S.; Sternefeld, W., editors. *Roots: Linguistics in search of its evidential base*. Berlin: Mouton de Gruyter; 2007. p. 77–96.
- Bresnan, J.; Cueni, A.; Nikitina, A.; Baayen, H. Predicting the dative alternation. In: Boume, G.; Kraemer, I.; Zwarts, J., editors. *Cognitive foundations of interpretation*. Amsterdam: Royal Netherlands Academy of Science; 2007. p. 69–34.
- Bresnan J, Ford M. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language*. 2010; 86:186–213.
- Bresnan J, Hay J. Gradient grammar: An effect of animacy on the syntax of *give* in New Zealand and American English. *Lingua*. 2008; 118:245–259.
- Brown M, Savova V, Gibson E. Syntax encodes information structure: Evidence from online reading comprehension. *Journal of Memory and Language*. 2011; 66:194–209.
- Chang F, Dell GS, Bock K. Becoming syntactic. *Psychological Review*. 2006; 113:234–272. [PubMed: 16637761]
- Clark, HH.; Haviland, S. Comprehension and the given-new contract. In: Freedle, R., editor. *Discourse production and comprehension*. Hillsdale, NJ: Erlbaum; 1977. p. 1–40.
- Clifton C Jr, Frazier L. Should given information come before new? Yes and no. *Memory and Cognition*. 2004; 32:886–895.
- Connine CM. Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language*. 1987; 26:527–528.
- Connine CM, Blasko DG, Hall M. Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language*. 1991; 30:234–250.
- Connine CM, Clifton C Jr. Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*. 1987; 13:291–299. [PubMed: 2953858]
- Davies, M. The Corpus of Contemporary American English (COCA): 410+ million words, 1990–present. 2008. Available online at <http://www.americancorpus.org>
- Ferreira VS, Dell GS. Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*. 2000; 40:296–340. [PubMed: 10888342]

- Ferreira VS, Hudson M. Saying “that” in dialogue: The influence of accessibility and social factors on syntactic production. *Language and Cognitive Processes*. 2011; 26:1736–1762.
- Fox Tree JE, Clark HH. Pronouncing “the” as “thee” to signal problems in speaking. *Cognition*. 1997; 62:151–167. [PubMed: 9141905]
- Frazier L, Clifton C Jr. Quantifiers undone: Reversing predictable speech errors in comprehension. *Language*. 2011; 87:158–171.
- Ganong WF. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*. 1980; 16:77–91.
- Gennari SP, MacDonald MC. Linking production and comprehension processes: The case of relative clauses. *Cognition*. 2009; 111:1–23. [PubMed: 19215912]
- Gerken L. The metrical basis for children’s subjectless sentences. *Journal of Memory and Language*. 1991; 30:431–451.
- Gibson E. Linguistic complexity: Locality of syntactic dependencies. *Cognition*. 1998; 68:1–76. [PubMed: 9775516]
- Haegeman L. Understood subjects in English diaries: On the relevance of theoretical syntax for the study of register variation. *Multilingua*. 1990; 9:157–199.
- Haegeman L. Register variation, truncation, and subject omission in English and in French. *English Language and Linguistics*. 1997; 1:233–270.
- Haegeman L. Subject omission in present-day written English: On the theoretical relevance of peripheral data. *Rivista di Grammatica Generativa*. 2007; 32:91–124.
- Haegeman L, Ihsane T. Subject ellipsis in embedded clauses in English. *English Language and Linguistics*. 1999; 3:117–145.
- Haegeman L, Ihsane T. Adult null subjects in the non-pro-drop languages: Two diary dialects. *Language Acquisition*. 2001; 9:329–346.
- Harris JA, Potts C. Perspective-shifting with appositives and expressives. *Linguistics and Philosophy*. 2009; 32:523–552.
- Hawkins, JA. A performance theory of order and constituency. Cambridge: Cambridge University Press; 1994.
- Hawkins, JA. Efficiency and complexity in grammars. Oxford: Oxford University Press; 2004.
- Hemforth, B.; Konieczny, L.; Scheepers, C.; Colonna, S.; Schimke, S.; Pynte, J. Language specific preferences in anaphor resolution: Exposure or Gricean maxims?. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*; 2010.
- Hinrichs L, Szmrecsanyi L. Recent changes in the function and frequency of Standard English genitive constructions: A multivariate analysis of tagged corpora. *English Language and Linguistics*. 2005; 11:437–474.
- Isenberg, D.; Walker, ECT.; Ryder, JM. A top-down effect in the identification of function words. Paper presented at the meeting of the Acoustical Society of America; Los Angeles, CA. 1980.
- Jaeger F. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*. 2010; 61:23–62. [PubMed: 20434141]
- Kaiser E, Trueswell JC. The role of discourse context in the processing of a flexible word-order language. *Cognition*. 2004; 94:113–147. [PubMed: 15582623]
- Levy, R. A noisy-channel model of rational human sentence comprehension under uncertain input. *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*; 2008a. p. 234–243.
- Levy R. Expectation-based syntactic comprehension. *Cognition*. 2008b; 106:1126–1177. [PubMed: 17662975]
- Levy, R. Integrating surprising and uncertain-input models in online sentence comprehension: Formal techniques and empirical results. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*; 2011.
- Levy R, Bicknell K, Slattery T, Rayner K. Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*. 2009; 106:21086–21090.

- Mack, J.; Fuerst, Y. English optional expletives and the pragmatics of judgments. *Proceedings of the 45rd Annual Meeting of the Chicago Linguistic Society*; (in press)
- Manning, CD. Probabilistic syntax. In: Bod, R.; Hay, J.; Jannedy, S., editors. *Probabilistic Linguistics*. Cambridge, MA: MIT Press; 2003. p. 289-341.
- Napoli DJ. Initial material deletion in English. *Glossa*. 1982; 16:85–111.
- Pickering MJ, Garrod S. Do people use language production to make predictions during language comprehension? *Trends in Cognitive Sciences*. 2007; 11:105–110. [PubMed: 17254833]
- R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2011. <http://www.R-project.org/>
- Rizzi, L. Early null subjects and root null subjects. In: Hoekstra, T.; Schwartz, B., editors. *Language Acquisition Studies in Generative Grammar*. Amsterdam: John Benjamins; 1994. p. 151-177.
- Rohde, D. Linger: A flexible platform for language processing experiments. 2003. <http://tedlab.mit.edu/~dr/Linger/>
- Rohde, H.; Ettlinger, M. Effects of pragmatic inference on phoneme identification. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*; 2010.
- Rosenbach A. Animacy versus weight as determinants of grammatical variation in English. *Language*. 2005; 81:613–644.
- Ruppenhofer, JK. Doctoral dissertation. University of California-Berkeley; 2004. The interaction of valence and information structure.
- Samuel AG. Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*. 1981; 110:474–494. [PubMed: 6459403]
- Samuel AG. The effect of lexical uniqueness on phonemic restoration. *Journal of Memory and Language*. 1987; 26:36–56.
- Samuel AG. Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General*. 1996; 125(1):28–51.
- Schmerling, SF. Subjectless sentences and the notion of surface structure. *Proceedings of the Annual Meeting of the Chicago Linguistic Society*; 1973. p. 577-586.
- Stoyneshka I, Fodor JD, Fernández EM. Phoneme restoration methods for investigating prosodic influences on syntactic processing. *Language and Cognitive Processes*. 2010; 25:1265–1293.
- Tabor W, Galantucci B, Richardson D. Effects of merely syntactic coherence on sentence processing. *Journal of Memory and Language*. 2004; 50:355–370.
- Temperley D. Minimization of dependency length in written English. *Cognition*. 2007; 105:300–333. [PubMed: 17074312]
- Thompson, SA. Information flow and dative shift in English discourse. In: Edmondson, JA.; Feagin, C.; Mühlhäusler, P., editors. *Development and diversity: Linguistic variation across time and space*. Dallas, Texas: SIL; 1990. p. 239-253.
- Thompson SA, Mulac A. The discourse conditions for the use of the complementizer *that* in conversational English. *Journal of Pragmatics*. 1991; 15:237–251.
- Thrasher, R. One way to say more by saying less: A study of so-called subjectless sentences. Tokyo: The Eihosha Ltd; 1977.
- Tily H, Gahl S, Arnon I, Snider N, Kothari A, Bresnan J. Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition*. 2009; 1:147–165.
- Van Alphen P, McQueen JM. The time-limited influence of sentential context on function word identification. *Journal of Experimental Psychology: Human Perception and Performance*. 2001; 27:1057–1071. [PubMed: 11642695]
- Warren RM. Perceptual restoration of missing speech sounds. *Science*. 1970; 167:392–393. [PubMed: 5409744]
- Wasow, T. Gradient data and gradient grammars. *Proceedings of the 43rd Annual Meeting of the Chicago Linguistic Society*; 2009. p. 255-271.
- Weir, A. English Language and Linguistics. Left-edge deletion in English and subject omission in diaries. (In press)
- Zwicky, A. Saying more with less. *Language Log* post. 2005. retrieved October 5, 2011 (<http://itre.cis.upenn.edu/~myl/language-log/archives/001995.html>)

Appendix: Study 2 Stimuli

-
1	Present	A: Joe and I are studying like crazy for the German exam.
B: Oh yeah?
A: It Seems to me/him like there are thousands of words to learn.
	Past	A: Joe and I studied like crazy for the German exam.
B: Oh yeah?
A: It Seemed to me/him like there were thousands of words to learn.
2	Present	A: Sorry I missed your call! I'm at the new Vietnamese restaurant with Molly.
B: How is it?
A: It Seems to me/her like the food is pretty Americanized.
	Past	A: Sorry I missed your call! I was at the new Vietnamese restaurant with Molly.
B: How was it?
A: It Seemed to me/her like the food was pretty Americanized.
3	Present	A: Christine and I are giving up on the project.
B: Why?
A: It Seems to me/her like there isn't any option.
	Past	A: Christine and I gave up on the project.
B: Why?
A: It Seemed to me/her like there wasn't any option.
4	Present	A: Tom and I are arguing about the article on stem cell research.
B: What's the problem?
A: It Seems to me/him like a lot of it is really biased.
	Past	A: Tom and I were arguing about the article on stem cell research.
B: What was the problem?
A: It Seemed to me/him like a lot of it was really biased.
5	Present	A: Some of us are thinking about betting on the game.
B: Oh yeah?
A: It Seems to me/Brad like we have a real shot at winning.
	Past	A: Some of us were thinking about betting on the game.
B: Oh yeah?
A: It Seemed to me/Brad like we had a real shot at winning.
6	Present	A: So, we don't have a theme for the fundraiser.
B: Why?
A: It Seems to me/Lauren like all of the options are kind of cliché.
	Past	A: So, we didn't have a theme for the fundraiser.
B: Why?
A: It Seemed to me/Lauren like all of the options were kind of cliché.
7	Present	A: My French class is terrible this week.
B: Why?
A: It Seems to me/the teacher like nobody understands the subjunctive.
	Past	A: My French class was terrible last week.
B: Why?
A: It Seemed to me/the teacher like nobody understood the subjunctive.

8. Present A: I'm calling the dealership about the weird smell coming from the car.
B: Oh yeah?
A: It Seems to me/the mechanic like it might be something serious.
Past A: I called the dealership about the weird smell coming from the car.
B: Oh yeah?
A: It Seemed to me/the mechanic like it might have been something serious.
9. Present A: Sam and I are reading the report on immigration reform.
B: And?
A: It Sounds to me/him like the bill doesn't have much of a chance.
Past A: Sam and I just read the report on immigration reform.
B: And?
A: It Sounded to me/him like the bill didn't have much of a chance.
10. Present A: Tina and I aren't sure whether to buy tickets for the play.
B: Oh yeah?
A: It Sounds to me/her like it's going to be pretty good.
Past A: Tina and I weren't sure whether to buy tickets for the play.
B: Oh yeah?
A: It Sounded to me/her like it was going to be pretty good.
11. Present A: Justin and I are reading about safety in Moscow.
B: And?
A: It Sounds to me/him like it's ok to walk around at night.
Past A: Justin and I were reading about safety in Moscow.
B: And?
A: It Sounded to me/him like it was ok to walk around at night.
12. Present A: Tracy and I are reading through the customer feedback on the new coffee flavors.
B: Mm hm?
A: It Sounds to me/her like people still like the house blend best.
Past A: Tracy and I read through the customer feedback on the new coffee flavors.
B: Mm hm?
A: It Sounded to me/her like people still liked the house blend best.
13. Present A: Things are pretty heated at union meetings this week.
B: Why's that?
A: It Sounds to me/the president like we won't be able to negotiate for another year.
Past A: Things were pretty heated at union meetings last week.
B: Why was that?
A: It Sounded to me/the president like we wouldn't be able to negotiate for another year.
14. Present A: I'm on the phone with the pediatrician's office.
B: And?
A: It Sounds to me/the nurse like we should bring Tommy in.
Past A: I was just on the phone with the pediatrician's office.
B: And?
A: It Sounded to me/the nurse like we should have brought Tommy in.
15. Present A: Today everyone's talking about why Bill got fired.

- B: Oh yeah?
- A: It Sounds to me/Diana like something weird's going on.
- Past A: Yesterday everyone was talking about why Bill got fired.
- B: Oh yeah?
- A: It Sounded to me/Diana like something weird was going on.
16. Present A: None of us can decide when to have the tournament.
- B: Oh yeah?
- A: It Sounds to me/Tim like the fall might be the best time.
- Past A: None of us could decide when to have the tournament.
- B: Oh yeah?
- A: It Sounded to me/Tim like the fall might have been the best time.
17. Present A: I'm going to Walmart tonight with Anna.
- B: Why?
- A: It Looks to me/her like they might have laptops on sale.
- Past A: I went to Walmart last night with Anna.
- B: Why?
- A: It Looked to me/her like they might have had laptops on sale.
18. Present A: Luke and I are going to the new gym on campus.
- B: Why?
- A: It Looks to me/him like they're offering some interesting classes.
- Past A: Luke and I went to the new gym on campus.
- B: Why?
- A: It Looked to me/him like they were offering some interesting classes.
19. Present A: Amy and I are really enjoying the Russian film festival.
- B: Really?
- A: It Looks to me/her like the organizers know their stuff.
- Past A: Amy and I really enjoyed the Russian film festival.
- B: Really?
- A: It Looked to me/her like the organizers knew their stuff.
20. Present A: Noah and I can't decide when to plant the tomatoes.
- B: Why?
- A: It Looks to me/him like the weather's not cooperating.
- Past A: Noah and I couldn't decide when to plant the tomatoes.
- B: Why?
- A: It Looked to me/him like the weather wasn't cooperating.
21. Present A: Several of us are reading through the fellowship applications today.
- B: How are they?
- A: It Looks to me/Martin like nobody has enough experience.
- Past A: Several of us read through the fellowship applications yesterday.
- B: How were they?
- A: It Looked to me/Martin like nobody had enough experience.
22. Present A: We're moving the party outside.
- B: Why?

- A: It Looks to me/Heather like it's going to be a big crowd.
- Past A: We moved the party outside.
- B: Why?
- A: It Looked to me/Heather like it was going to be a big crowd.
23. Present A: My parents and I don't think we're going to make it to the wedding.
- B: Why not?
- A: It Looks to me/Dad like our flight might be delayed.
- Past A: My parents and I didn't think we were going to make it to the wedding.
- B: Why not?
- A: It Looked to me/Dad like our flight might have been delayed.
24. Present A: My roommates and I are taking a class on medical administration.
- B: How is it?
- A: It Looks to me/Andy like it's a pretty cool career path.
- Past A: My roommates and I took a class on medical administration.
- B: How was it?
- A: It Looked to me/Andy like it was a pretty cool career path.
-

Highlights

- Two studies tested pragmatic effects on the expletive/zero subject alternation.
- A corpus study showed that zero sentences tend to express immediate judgments.
- A speech restoration study found more zero responses in immediate contexts.
- Subtle pragmatic age preferences affect what listeners report hearing.

Table 1

Frequency of expletive and zero sentences with present- and past-tense verbs in the COCA

	Expletive Subject	Zero Subject	Expletive : Zero Ratio
Present-Tense Verb	677	159	4.3:1
Past-Tense Verb	137	16	8.6:1

Table 2

Frequency of expletive and zero sentences by verb in the COCA

	Expletive Subject	Zero Subject	Expletive : Zero Ratio
Verb: <i>Seem</i>	302	33	9.2:1
Verb: <i>Look</i>	396	93	4.3:1
Verb: <i>Sound</i>	116	49	2.4:1

Table 3

Parameters of the selected model of the corpus data

	Coefficient Estimate	Standard Error	Z Value	P value
(Intercept)	1.7202	0.1449	11.875	< 0.0001
Tense: Present	−0.2823	0.1412	−2.000	0.0455
Verb: <i>Seem</i>	0.6778	0.1408	4.814	< 0.0001
Verb: <i>Sound</i>	−0.6217	0.1352	−4.599	< 0.0001

Table 4

Mean (SD) length in seconds of distorted segments across conditions

	Tense: Present	Tense: Past
Person: First	.086 (.016)	.086 (.012)
Person: Third	.081 (.012)	.081 (.013)

Table 5

Speech restoration study, spoken responses

	<i>present, 1st-person</i>	<i>present, 3rd-person</i>	<i>past, 1st-person</i>	<i>past, 3rd-person</i>
# <i>it</i> responses	179	184	185	199
# zero responses	107	91	81	67
% <i>it</i> restoration	62.6%	66.9%	69.5%	74.8%

Table 6

Parameters of the selected model of the speech restoration data (spoken responses)

	Coefficient Estimate	Standard Error	Z Value	P value
(Intercept)	1.40316	0.33060	4.244	< 0.0001
Tense: Present	−0.29088	0.14319	−2.032	0.04220
Person: First	−0.18685	0.09899	−1.888	0.05908
Verb: <i>Seem</i>	0.41322	0.14737	2.804	0.00505
Verb: <i>Sound</i>	−0.29612	0.16950	−1.747	0.08064

Table 7

Parameters of the selected model of the “extreme” subset of the speech restoration data (spoken responses)

	Coefficient Estimate	Standard Error	Z Value	P value
(Intercept)	1.4728	0.3614	4.076	< 0.0001
Immediacy: Immediate	−0.5219	0.1545	−3.379	0.000728
Verb: <i>Seem</i>	0.6904	0.2494	2.768	0.005641
Verb: <i>Sound</i>	−0.6912	0.1795	−3.851	0.000118

Table 8

Speech restoration study, written responses

	<i>present, 1st-person</i>	<i>present, 3rd-person</i>	<i>past, 1st-person</i>	<i>past, 3rd-person</i>
# <i>it</i> responses	187	190	182	193
# zero responses	75	66	54	48
% <i>it</i> restoration	71.4%	74.2%	77.1%	80.1%