Avoiding interpretation ambiguities with phonetic cues*

Simon Todd

Qualifying Paper 1
Department of Linguistics, Stanford University

December 31, 2014

1 Introduction

Imagine you are talking to your sister on a phone with a noisy connection. She has recently got a cat and is telling you about life with it, and you hear her utter (1) (where – indicates noise):

\[ I \text{ want} - \text{ give} - \text{ my cat} - \text{ a friend}. \] (1)

In spite of the noise, you’re able to understand what you think your sister is trying to tell you: that she wants to get another cat (i.e. you interpret her utterance as \textit{I want to give my cat a friend}). However, when you then ask her what breed she wants to get, she is confused, because she has decided that she can’t manage owning a cat; what she actually said was \textit{I want to give my cat to a friend}.

Now imagine you call your brother and tell him about your conversation. You want to report what your sister told you about her cat, but saying \textit{she wants to give a friend her cat} seems strange. While there are many other ways in which you could convey the message, for whatever reason you use the same structure that your sister did: \textit{she wants to give her cat to a friend}. However, you know that losing the \textit{to} over a noisy connection could force your brother to the wrong interpretation, as happened to you. How can you ensure that he arrives at the correct interpretation, without changing the structure you use to convey it?

*This paper uses data from the Origins of New Zealand English (ONZE) project at the University of Canterbury. The Mobile Unit data was collected by the Mobile Disc recording unit of the NZ Broadcasting Service, the Intermediate Corpus data was collected by Rosemary Goodyear, Lesley Evans and members of the ONZE team. The work done by members of the ONZE team in preparing the data, making transcripts and obtaining background information is gratefully acknowledged. I am very grateful to Daniel Bürkle, Kirsty Thompson, Liam Walsh and Alia Hope-Wilson for their help with extracting, realigning and coding the dative constructions, to Jen Hay for facilitating that help, and to Anette Rosenbach, Sali Tagliamonte and Joan Bresnan for their guidance in developing coding criteria and assistance with correcting coding errors. I’d like to thank my QP committee – Joan Bresnan, Dan Jurafsky, and John R. Rickford – for helpful discussions on theoretical and methodological matters. This material is based upon work supported by the National Science Foundation under Grant No. BCS-1025602.
This paper explores the hypothesis that such interpretation ambiguities are avoided by phonetic means; namely, that the phonetic cues to critical words that exclusively support the intended interpretation (like to) are stronger when the intended interpretation could not be confidently inferred without them. This hypothesis predicts, for example, that the phonetic realization of to in your reproduction of (1) for your brother should be stronger (in some way) than in a version of (1) with my daughter in place of a friend, where the alternative interpretation is less (or perhaps no longer) plausible. In this paper, I develop an interpretability model for quantifying the extent to which an intended interpretation is able to be inferred without the critical word to in English dative constructions, and use it with a corpus of spontaneous speech to test the prediction that phonetic cues to critical words are strengthened in low-interpretability utterances.

The rest of this paper is organized as follows: Section 2 reviews relevant concepts from the literature, introducing the notion of information-based communication and the dative alternation, showing how previous studies have indicated that they may be linked, and indicating the way in which the present study attempts to extend that link. Section 3 introduces the data that are used for this study, describing the corpus from which dative constructions were obtained and the annotations that were made of these constructions. Section 4 describes the formulation of the interpretability model through models for syntactic alternant choice and thematic role assignment, and indicates how its relationship to the strength of phonetic cues to the critical word to is assessed. Section 5 presents the results of this assessment and Section 6 scrutinizes these results in light of alternative explanations and data limitations. Section 7 discusses potential implications of the results and points to further work that can be done to clarify these implications, and Section 8 concludes.

2 Background

2.1 Information-based communication

The notion that critical words should be strengthened in low-interpretability utterances has its roots in information theory (Shannon, 1948). Under this view, communication comprises three main parts: the encoding of a message in a signal (e.g. a speaker converting a concept to a string of words); the transmission of that signal through a potentially noisy channel (e.g. the physical acts of the speaker saying those words and a listener hearing them); and the decoding of the signal to retrieve the message (e.g. the listener comprehending the utterance). The communicative act is successful if the decoded message matches the original (encoded) message. Two opposing forces act on the design of a signal: efficiency, which pressures the signal to be short and simple so that the communicative act is not effortful or time-consuming; and robustness, which pressures the signal to be full and elaborate so that the communicative act is successful. Optimal signals balance these forces by minimizing redundant material (which is largely uninformative of the message, often because it is predictable from context) and maximizing informative material (which is unpredictable or otherwise highly important for successful decoding). In this way,
optimal communication involves the transfer of information at a constant rate.

A number of authors have argued that natural language production approximates that expected from an optimal communication system (Aylett and Turk, 2004, 2006; Genzel and Charniak, 2002; Jaeger, 2010; Kurumada and Jaeger, 2013; Levy and Jaeger, 2007; van Son and Pols, 2003). In speech, Aylett and Turk (2004, 2006) argue that near-optimality arises through the reduction (in terms of duration and vowel quality) of syllables that are predictable from syllabic, word and discourse context, a phenomenon which has been separately confirmed (for words) by Bell et al. (2003, 2009). Furthermore, van Son and Pols (2003) present evidence for the phonetic strengthening of segments which most strongly disambiguate potential intended words, suggesting that near-optimality of spoken communication may also arise through consideration of what is informative for the listener.

The notion that the speaker takes account of the listener’s information needs does not imply that the listener dictates the linguistic choices of the speaker; rather, the speaker must balance the listener’s need for robust, informative cues to the message against her own needs for efficient formulation and transfer of the signal. This is why, for example, speakers don’t avoid ambiguous syntactic structures (Arnold et al., 2004): context promotes disambiguation, making the use of a more complex, unambiguous structure redundant and inefficient (Piantadosi et al., 2012). The inferential system works faster than the production system (MacDonald, 2013); so, provided the message can be confidently inferred from the signal, requiring the listener to make this inference is more optimal than requiring the producer to explicate it. However, small-scale production operations such as the realization of an optional monosyllabic case-marker (Kurumada and Jaeger, 2013) and the modulation of prosody (Snedeker and Trueswell, 2003) are arguably not very effortful and may be used to aid the comprehension of listeners when context supports alternative interpretations of an utterance. I will refer to such usage as audience design.

2.2 The dative alternation

The dative alternation refers to the two possible ways of ordering the NP arguments of certain ditransitive verbs, which have the thematic roles of theme and recipient. In the double-object dative (a), the recipient (my daughter) appears before the theme (my cat), while the reverse is true in the prepositional dative (b) (which also separates the two arguments by the preposition to):

(a) ... give my daughter my cat.

(b) ... give my cat to my daughter.

Many factors have been proposed to be relevant to the choice of dative alternant (see reviews in Bresnan et al., 2007; Bresnan and Ford, 2010). Using a large corpus of spontaneous (American English) speech, Bresnan et al. (2007) and Bresnan and Ford (2010) showed that these factors may be combined in a probabilistic model to predict the choice of alternant with extremely high (95%) accuracy. Specifically, the model revealed constraints that privilege pronominal, animate, definite and short arguments, promoting their ordering before non-pronominal,
inanimate, indefinite and long arguments. These constraints exist in parallel with each other (and others) and thus may reinforce or interfere with each other on any given utterance, depending on its properties. In this way, the model yields a continuum of probabilities with which one alternant will be chosen over the other in different situations.

A series of experiments (Bresnan, 2007; Bresnan and Ford, 2010) have shown that language-users not only shape their production choices following these probabilities, but also have access to them in online and offline perception and judgment tasks. As the likelihood of a given dative being produced in the prepositional form increases, readers rate the prepositional alternant as more natural and respond to the word *to* faster in sequential processing. Thus, alternant choice probabilities are cognitively real and are used to build expectations that aid comprehension.

### 2.3 Predictability and phonetic realizations in datives

The cognitive reality of syntactic probabilities implies that, if natural language forms a near-optimal communication system, speakers ought to pronounce elements of predictable (high-probability) constructions in a relatively more reduced fashion compared to in unpredictable (low-probability) constructions (Gahl and Garnsey, 2004; Gahl et al., 2006). Tily et al. (2009) found that this was the case for the dative alternation: in the same corpus of spontaneous (American English) speech used by Bresnan et al. (2007) to fit the syntactic alternant choice model, instances of *to* that occurred in datives that were highly expected to take the prepositional form were shorter than those that occurred in datives that were unexpected to take the prepositional form.

However, this result has been called into question. Kuperman and Bresnan (2012) revisited Tily et al.’s (2009) data, applying a more stringent statistical analysis, and failed to find a difference in the duration of *to* according to alternant choice probability. Drawing on a particular instantiation of the theory of information-based communication (Uniform Information Density: Levy and Jaeger, 2007; Jaeger, 2010), they argue that this is not unexpected, as the unfolding of the entire construction is perfectly predictable for the speaker once she has committed to a particular alternant, which must occur before the first argument (and thus *to*) is uttered. This suggests that Tily et al.’s (2009) result reflected a non-causal correlation and calls for a refinement of the predictor variable and its theoretical motivation.

### 2.4 Proposal: interpretability and phonetic realizations in datives

Kuperman and Bresnan (2012) suggest that the mechanistic reason behind their failure to find the effect of alternant choice probability on *to*-duration reported by Tily et al. (2009) is due to differences in the statistical analysis.

---

1. They did, however, find probabilistic effects on the duration of the verb and that of the word after the verb: both words were shorter in higher-probability alternants. They argue that the effects on the first word after the verb reflect information-based processes (as this word commits the speaker to a given alternant, which may be more or less predictable), while the effects on the verb itself reflect planning processes (as the speaker is diverting cognitive resources to plan the construction whilst uttering the verb, and planning of a high-probability alternant induces less planning effort than that of a low-probability alternant).
Specifically, they show that it is log-transforming the duration variable\(^2\) which removes the effect. Nevertheless, to-duration exhibits substantial variation in their dataset, ranging from 20 to 400ms (with a mean duration of 108ms and standard deviation of 73ms; p. 598). This implies that a suitable function of alternant choice probability may be able to ‘reverse’ the effects of the transformation of duration; thus, a new predictor variable, non-linearly constructed from alternant choice probability, may demonstrate a relationship with (transformed) to-duration.

What form should such a predictor take? It is clear from Kuperman and Bresnan’s (2012) argument that it cannot simply reflect the predictability of the alternant from the speaker’s point of view, as the speaker is subject to no uncertainty about the alternant once she has committed to it. I propose instead that it should reflect the interpretability from the listener’s point of view – that is, the extent to which the listener would be able to correctly infer the speaker’s intended message without the additional cue provided by the critical word to. As discussed in Section 2.1, this is a measure of informativity which ought to be relevant to simple processes in speech if natural language approximates an optimal communication system; and, as I will show in Section 4.1, it is constructed in part from the syntactic alternant choice probability.

The cognitive assumption underlying the use of interpretability for this purpose is that the speaker has a model of the inferential task faced by the listener in comprehension, which reflects her own experience of being a listener in other situations. During production, the speaker exhibits audience design and references\(^3\) this model for areas where the listener may struggle to retrieve her intended interpretation, which she then strengthens in small-scale ways (e.g. lengthening words) so as to increase the probability that the communicative act will be successful. In this way, she is hypothesized to lengthen the critical word to (or its component phones) in prepositional datives when failure to perceive the to would lead the listener to infer a double-object interpretation, so as to increase the robustness of the information transfer.

While this hypothesis is similar in spirit to the notion that speakers use prosody to avoid syntactic attachment ambiguities (Snedeker and Trueswell, 2003), it extends that notion by assuming a quantitative model of interpretability. It also offers a novel extension of the notion that natural language forms a near-optimal communication system, as similar studies explicitly taking this stance have either explored acoustic-phonetic gradience as it relates to the (speaker-based) predictability of syntactic structures (Kuperman and Bresnan, 2012; Tily et al., 2009) or have only considered non-phonetic, non-gradient effects of (listener-based) interpretability (Kurumada and Jaeger, 2013).

\(^2\)Kuperman and Bresnan (2012) state that the log-transformation is motivated by a desire to “attenuate skewness in the distribution of acoustic durations and... render the distribution closer to normality, as required for the improved accuracy of parameters in regression models” (pp. 598-599). While this is common in the psycholinguistic literature, the stated motivation is technically incorrect: an appropriate transformation should be chosen to normalize the residuals of a model, not the raw data (see note 27). Kuperman (p.c.) states that the log-transformation did yield a more normal distribution of residuals, but it is not clear whether it was the most appropriate transformation for that purpose, as no other transformations are reported to have been tried.

\(^3\)Jaeger (2013) notes that aspects of audience design may be learned through experience and thus not require the active reference to a model of the listener’s comprehension process. I have chosen to discuss audience design here as though it were an active process simply for expository purposes.
3 Data

3.1 Data source

The data for this study is taken from the ONZE corpus, a corpus of recordings of people born in New Zealand from the 1850s to the 1980s (Gordon et al., 2007). In total, the corpus contains approximately 1.6 million words spanning 113 hours of speech from 650 speakers. It is split into three sub-corpora as follows:

The Mobile Unit contains recordings from approximately 100 speakers born between 1851 and 1910, recorded between 1946 and 1948. The recordings are individual and group interviews with varying levels of formality, collected for broadcast on national radio.

The Intermediate Archives contain recordings from approximately 140 speakers born between 1890 and 1930, recorded in the early 1990s. The recordings are individual interviews, collected for oral history research or for broadcast on national radio.

The Canterbury Corpus contains approximately 420 speakers born between 1926 and 1987, recorded between 1994 and 2008. The speakers are roughly balanced with respect to gender, age, and social class. The recordings are individual sociolinguistic interviews conducted by undergraduates, featuring casual conversation.

3.2 Data preparation

3.2.1 Data extraction and exclusion

Dative constructions were harvested from the corpus by searching for instances of the 38 dative verbs identified as alternating in the Switchboard Corpus by Bresnan et al. (2007), in any form (tense, aspect, person, etc.). These verbs are listed in Table 1. Following Bresnan et al. (2007), the results of this search were manually filtered so that only active dative uses of these verbs with two overt nominal objects which were judged capable of alternating between a single double-object and a single to form were included. This filtering process removed constructions with CP arguments (tell you that I love you, give you what you want), benefactive (make him some tea) and locative constructions (take him to Paris), transitive or elliptical uses (pay him), passive constructions (John was given a dog), rate constructions (pay me five dollars an hour), and fixed expressions (tell you what, do it to him). The search and filter process yielded 1613 constructions, of which 1375 were double-object datives and 238 were prepositional datives; this represents a significantly smaller proportion of prepositional datives than what has previously been reported for US English (Bresnan et al., 2007) (NZ: 238/1613, US: 501/2360; $\chi^2 = 26.09, p = 3.25 \times 10^{-7}$).

---

4 A questionable case was judged capable of alternating if a construction with the same general structure as the alternant not used was found on at least 10 unique English-based sites indexed by Google.

5 Benefactives were excluded because the double-object construction in these cases alternates with a for-construction rather than a to-construction; e.g. make him some tea ~ make some tea for him.

6 Rate constructions were excluded because it is not always clear whether the time-domain is part of the theme or a modifier of the verb, meaning that such constructions have in principle two plausible to forms; e.g. giving her a hundred dollars a week ~ giving a hundred dollars a week to her OR giving a hundred dollars to her a week.
Table 1: Alternating dative verbs extracted from ONZE.

<table>
<thead>
<tr>
<th>afford</th>
<th>allot</th>
<th>allow</th>
<th>assign</th>
<th>award</th>
<th>bet</th>
<th>bring</th>
<th>cause</th>
</tr>
</thead>
<tbody>
<tr>
<td>charge</td>
<td>cost</td>
<td>deny</td>
<td>do</td>
<td>feed</td>
<td>flip</td>
<td>float</td>
<td>give</td>
</tr>
<tr>
<td>hand</td>
<td>leave</td>
<td>lend</td>
<td>loan</td>
<td>mail</td>
<td>make</td>
<td>offer</td>
<td>owe</td>
</tr>
<tr>
<td>pay</td>
<td>promise</td>
<td>quote</td>
<td>read</td>
<td>sell</td>
<td>send</td>
<td>serve</td>
<td>show</td>
</tr>
<tr>
<td>swap</td>
<td>take</td>
<td>teach</td>
<td>tell</td>
<td>wish</td>
<td>write</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2.2 Data realignment

Because the hypothesis under investigation here concerns the phonetic realization of syntactic constructions, it was necessary to obtain accurate time-alignments for the corpus data. The time-alignments included in the corpus, which were automatically generated using the HTK toolkit (Young, 1994), were judged to be insufficiently accurate; thus, word-level alignments were manually adjusted for all words in each construction. This manual adjustment was done on the basis of visually-identifiable spectrographic transitions and auditory judgments. Additionally, instances of the word to which were of interest for the present study were hand-aligned at the segment level, with the segmental boundary between /t/ and /u/ placed so as to coincide with spectrographic and waveform cues to the onset of voicing⁷.

3.2.3 Data coding

The data were coded for factors which have been found to be significant predictors of alternant choice⁸ (Bresnan et al., 2007; Bresnan and Hay, 2008; Bresnan and Ford, 2010) and acoustic reduction (Jurafsky et al., 2002; Bell et al., 2009). For every construction, the recipient and theme were manually identified, as were their heads. Most of the coding was then conducted on the basis of the properties of these heads⁹.

Pronominality. The theme and recipient were coded with a binary variable representing the pronominality of their (syntactic) head. Following Bresnan and Ford (2010), the pronominal class was defined to include instances of personal pronouns (him, you, it), demonstrative pronouns (that), and reflexive pronouns (himself). Everything else, including indefinite pronouns (someone, everything), was defined to belong to the nonpronominal class. The coding of pronominality was automated via lookup of a list of the members of the closed pronominal class, and later

---

⁷The primary spectrographical cue used was the onset of the F0 voicing bar and the primary waveform cue used was the onset of regular periodic waves. Auditory checks confirmed that these cues aligned with an impressionistic perceptual separation of the phones.

⁸Discourse accessibility, or givenness, was not included here for practical reasons, even though it has been found to be a significant predictor of alternant choice in some studies (Bresnan et al., 2007; Bresnan and Hay, 2008). Coding for givenness would have required extensive manual annotation based on an awareness of the discourse context in which each construction occurred. It is instead assumed here that definiteness may act as proxy for givenness, as given entities are primarily referred to with definite expressions. For example, when referring to a car that has been previously established in the discourse, it would be highly unusual to use the indefinite a car instead of the definite the car. This decision is supported by the fact that givenness did not emerge as significant in Bresnan and Ford’s (2010) revised model.

⁹The syntactic heads were identified in this way, as they formed the basis of the coding that could be automated. Coding some properties, such as animacy, required referring to the semantic heads; however, as such coding was conducted manually, the identification of semantic heads (in cases where they differed from syntactic heads) was left to be conducted during the coding process. Semantic and syntactic heads were usually identical, differing only in a small number of cases such as the partitive tell one of his mates a secret.
hand-verified. The distribution of pronominalities across dative types is shown in Table 2; as can be seen, recipients tend to be pronominal while themes tend to be non-pronominal, and pronominal recipients and non-pronominal themes both favor the double-object (NP-NP) order.

**Animacy.** The theme and recipient were hand-coded with a binary variable representing the animacy of the referent of their (semantic) head. Humans and members of the animal kingdom were coded as *animate*, while all others (including organizations\(^{10}\)) were coded as *inanimate*. Table 3 shows the distribution of animacies across dative types, in which animate recipients and inanimate themes both favor the double-object order. It can also be seen that recipients are likely to be animate and themes are likely to be inanimate.

**Definiteness.** The theme and recipient were coded with a 3-valued variable representing their definiteness. *Definite proper nouns* were defined to be titles or the names of people, places or organizations, transcribed as beginning with a capital letter. *Definite* NPs were defined to be those that failed the *existential there be* test\(^{11}\); such NPs are typically preceded by a member of a closed class of determiners (*the, that, every*) or are pronouns or bare plurals (Garretson, 2004). *Indefinite* NPs were defined to be those that belong to neither of the other two classes. The coding of definiteness was automated via checking capitalization, plurality\(^{12}\), and pronominality\(^{13}\), and looking up the first word of the NP in the class of definite determiners provided by Garretson (2004). The automatically-assigned classes were later hand-verified. The distribution of definitenesses across dative types is shown in Table 4; definite recipients and indefinite themes both favor the double-object order, and the definite feature is associated more with recipients than with themes.

**Number.** The theme and recipient were coded with a binary variable representing whether they are *singular*

### Table 2: Pronominality of recipient and theme across dative types.

<table>
<thead>
<tr>
<th></th>
<th>Recipient</th>
<th></th>
<th>Theme</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pronominal</td>
<td>Non-pronominal</td>
<td>Pronominal</td>
<td>Non-pronominal</td>
</tr>
<tr>
<td>NP-NP</td>
<td>1247</td>
<td>128</td>
<td>110</td>
<td>1265</td>
</tr>
<tr>
<td>NP-PP</td>
<td>71</td>
<td>167</td>
<td>122</td>
<td>116</td>
</tr>
</tbody>
</table>

### Table 3: Animacy of recipient and theme across dative types.

<table>
<thead>
<tr>
<th></th>
<th>Recipient</th>
<th></th>
<th>Theme</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Animate</td>
<td>Inanimate</td>
<td>Animate</td>
<td>Inanimate</td>
</tr>
<tr>
<td>NP-NP</td>
<td>1318</td>
<td>57</td>
<td>15</td>
<td>1360</td>
</tr>
<tr>
<td>NP-PP</td>
<td>177</td>
<td>61</td>
<td>6</td>
<td>232</td>
</tr>
</tbody>
</table>

\(^{10}\)Following Garretson (2004), organizations were required to have a collective voice; groups of humans without a collective voice were coded as *animate*.

\(^{11}\)An NP fails the test if it is not acceptable or lacks an existential reading in a context such as *there is/are _____ outside*.

\(^{12}\)As outlined in the *Number* description.

\(^{13}\)As outlined in the *Pronominality* description.
Table 4: Definiteness of recipient and theme across dative types.

<table>
<thead>
<tr>
<th></th>
<th>Recipient</th>
<th></th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Indefinite</td>
<td>Definite</td>
<td>Definite-PN</td>
</tr>
<tr>
<td>NP-NP</td>
<td>29</td>
<td>1310</td>
<td>36</td>
</tr>
<tr>
<td>NP-PP</td>
<td>42</td>
<td>172</td>
<td>24</td>
</tr>
</tbody>
</table>

or plural. Plural NPs were defined as those with a plural quantifier (two, some, every) and those whose head has formal plural marking (-s, -es), including irregular plural pronouns (us, them). Bare mass nouns (water, cheese) were also coded as plural. All other NPs were coded as singular. The coding of number was automated via lookup of quantifiers and simple morphological parsing, then hand-verified. Hand-verification was particularly important for nouns with null plural marking, such as fish. Table 5 shows the distribution of numbers across dative types and reveals that number information creates no apparent bias for dative alternant choice (recipient: $\chi^2 = 2.36, p = 0.12$; theme: $\chi^2 = 2.38, p = 0.12$). It also shows that more recipients than themes are plural and more themes than recipients are singular.

Syntactic complexity. The theme and recipient were automatically coded for the number of words they contain based on whitespace separation, as a proxy for syntactic complexity (Shih and Grafmiller, 2011). To facilitate this process, transcript annotations, repairs (a ma–woman), repeats (a a man), fillers (um, uh) and discourse markers (like, sort of) were removed, characters such as $ were converted to full words ($10 becomes ten dollars), and hyphenated forms were split (bear-hug becomes bear hug). Additionally, contractions were split (don’t becomes do n’t) to be counted as two words, numbers were spelled out (1992 becomes nineteen ninety two) to be counted as multiple words, and abbreviations were joined (E P A becomes EPA) to be counted as a single word. The relative complexity of the two arguments was then calculated using equation (2) (Bresnan and Ford, 2010). The distributions of relative complexity across dative types are shown in Figure 1; as can be seen, when the recipient is shorter than the theme (i.e. when relative complexity is negative), the double-object alternant is favored. It is also apparent from the frequency values that it is typical for themes to be longer than recipients.

$$\text{Relative complexity} = \log(\text{recipient length}) - \log(\text{theme length}) \quad (2)$$

Table 5: Number of recipient and theme across dative types.

<table>
<thead>
<tr>
<th></th>
<th>Recipient</th>
<th></th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Plural</td>
<td>Singular</td>
<td></td>
</tr>
<tr>
<td>NP-NP</td>
<td>369</td>
<td>1006</td>
<td></td>
</tr>
<tr>
<td>NP-PP</td>
<td>52</td>
<td>186</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: Relative complexity across dative types.

**Structural persistence.** Each construction was automatically coded with the type (prepositional or double-object) of the most recent dative previously said by the same speaker (or with an X if the speaker had not used any dative previously), as well as the time elapsed between the end of that construction and the beginning of the present construction. To capture the importance of recency in structural parallelism effects (Szmrecsanyi, 2005), these two variables were combined in a single variable with value equal to the type of the dative used most recently if it occurred within the previous minute, or to X if no dative was used within the previous minute. Table 6 shows the distribution of most recent dative types across current dative types and suggests that the tendency to produce a prepositional dative may be increased after a recent prepositional dative.

**Speaker year of birth.** The year of birth of the speaker was extracted from ONZE for each construction. Under the apparent time hypothesis, this information can be used to study how the alternation has changed over time; data from speakers born long ago can be used to represent the linguistic state of affairs long ago, even when that data was not collected long ago (Cukor-Avila and Bailey, 2013).

**Verb sense.** For each construction, the lemma of the verb was extracted and manually classified as belonging to one of 5 broad semantic classes (Bresnan et al., 2007). The classes make distinctions between transfer of an object (*give her a car*), future transfer of an object (*offer her a car*), communication of information (*tell them a story*), prevention of possession of an object (*deny her a car*), and abstract (*give me a headache*). The pair consisting of a verb and its semantic class will be referred to as verb sense. There are 40 different verb senses in the data, which occur with various frequencies; the most frequent verb senses are *give-transfer* (371 instances), *give-abstract* (337

Table 6: Type of most recent dative within 1 minute across current dative types.

<table>
<thead>
<tr>
<th></th>
<th>Recent NP-NP</th>
<th>Recent NP-PP</th>
<th>No recent dative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current NP-NP</td>
<td>320</td>
<td>49</td>
<td>1006</td>
</tr>
<tr>
<td>Current NP-PP</td>
<td>42</td>
<td>28</td>
<td>168</td>
</tr>
</tbody>
</table>
instances) and *tell-communication* (232 instances).

**Speech rate.** Speech rate was desired as a normalization factor in the duration of *to* in prepositional datives, so that speech samples at different rates could be compared to each other. While global speech rate information was available for each speaker, averaged across the entire recording, this was not used as it falsely assumes that a speaker will always talk at the same rate. Rather, a local measure of speech rate was developed for prepositional datives by dividing the total number of syllables\(^{14}\) in the verb, theme and recipient (i.e. the number of syllables in the construction excluding the word *to*) by the total duration of these components, as in equation (3):

\[
\text{Local speech rate} = \frac{\# \text{ syllables in verb} + \# \text{ syllables in theme} + \# \text{ syllables in recipient}}{\text{duration of verb} + \text{duration of theme} + \text{duration of recipient}}
\]  

**Conditional probabilities.** For the prepositional datives, it was desired to include as a control the predictability of *to* given the words around it, as *to* is expected to be more reduced when it is more predictable. Two such measures were calculated for each construction: the *forward-bigram probability* of *to* given the last word of the theme, and the *backward-bigram probability* of *to* given the first word of the recipient. These measures were calculated on the basis of (directional) co-occurrence frequencies in the Web1T corpus\(^{15}\) (Brants and Franz, 2006); the conditional probability is given by the frequency of the bigram (pair of words) in the corpus, divided by the frequency of the last word of the theme (in the forward case) or of the first word of the recipient (in the backward case), as shown by equations (4) and (5):

\[
P(to|\text{theme lastword}) = \frac{\text{count}(\text{theme lastword, to})}{\text{count}(\text{theme lastword})} = \frac{\text{count}(\text{theme lastword, to})}{\sum_{\text{anyword}} \text{count}(\text{theme lastword, anyword})}
\]  

\[
P(to|\text{rec firstword}) = \frac{\text{count}(\text{to, rec firstword})}{\text{count}(\text{rec firstword})} = \frac{\text{count}(\text{to, rec firstword})}{\sum_{\text{anyword}} \text{count}(\text{anyword, rec firstword})}
\]  

**Phonological environment.** For prepositional datives, the identity of the phoneme immediately preceding *to* was automatically obtained from the phonological form of the last word of the theme, as given in CELEX (Baayen et al., 1995). Cases not included in CELEX were coded by hand. The realization of /t/ in *to* is expected to vary as a function of this phoneme.

---

\(^{14}\)Syllable numbers for English common nouns were extracted from CELEX (Baayen et al., 1995), and syllable numbers for proper nouns and non-English words were manually coded.

\(^{15}\)The Web1T corpus was chosen primarily because of its large size. While it is true that Web1T is not representative of the same type of language as ONZE is – because it pools usage across numerous varieties of English, because it is based on written data, and because the distribution of topics and styles is likely to be very different than that seen in interviews – it is assumed that these matters should not heavily influence the use of the word *to*, which is what the predictability measure is based on. Additionally, it is unreasonable to estimate collocation frequencies on the basis of ONZE data because the relatively small size of ONZE and the fact that it is the source of the datives together mean that the frequencies of collocations occurring in the datives would be biased through inflation.
4 Methods and models

4.1 Interpretability model

The idea underlying the hypothesis explored here is that a speaker should strengthen the word to in a prepositional dative – a strong cue to the intended interpretation of the utterance – when other aspects of the construction (i.e. properties of the verb and its arguments) do not give sufficiently strong cues to the intended interpretation. The interpretability model is thus construed as a model of the probability that a listener hearing a verb-NP-NP sequence (representing a double-object dative or a prepositional dative with a sufficiently weak to that it is lost in the noise of transmission) will assign the thematic roles (theme and recipient) to the NP arguments of the verb in the same way that the speaker did.

To put it another way, the task of the listener in such a situation is to decide which of the two NP arguments of the verb is the theme and which is the recipient, given the verb, the arguments, and the order in which the arguments occur. To make this decision, the listener must compare how expected the observed utterance is under each of the two possible assignments of thematic roles to arguments. For each assignment, this degree of consistency with the observation is formed by considering how likely it is for the thematic roles to be assigned in that way simply by virtue of the intrinsic properties of the arguments and how likely it is that, given those thematic role assignments, the arguments will be arranged in the order in which they appear in the construction. The weighing of one hypothesized thematic role assignment against the other yields a probability that the speaker intended for the construction to be interpreted one way as opposed to another.

The interpretability model captures the probability just described with respect to the actual interpretation intended by the speaker (as opposed to the alternative, unintended interpretation). It does this through decomposing the problem in the way just described, using Bayes’ theorem, as in (6) (see Appendix A for a derivation):

\[
P(\Theta = \theta | v, w, o) = \frac{P(O = o | v, w, \theta) \times P(\Theta = \theta | v, w)}{\sum_{\theta'} [P(O = o | v, w, \theta') \times P(\Theta = \theta' | v, w)]} \tag{6}
\]

where \(v\) represents the verb sense, \(w\) represents properties of the arguments (e.g. animacy, pronominality, definiteness), \(\theta\) represents the intended assignment of thematic roles, \(\Theta\) is a random variable over possible assignments of thematic roles, \(o\) represents the actual order in which the arguments are produced, and \(O\) is a random variable over possible orders in which the arguments could be produced.

The terms in (6) are obtained from a syntactic alternant choice model and a thematic role assignment model, which will be described in detail in Sections 4.2 and 4.3 respectively. \(P(O = o | v, w, \theta)\) is the probability of ordering the arguments in the observed order given their properties, the thematic roles that are assigned to them and the verb sense, obtained from the syntactic alternant choice model via (7). Similarly, \(P(\Theta = \theta | v, w)\) is the prior probability that the thematic roles are assigned as the speaker intended purely due to intrinsic properties of the
arguments\textsuperscript{16}, obtained from the thematic role assignment model via (8). The denominator in (6) is a sum over possible thematic role assignments which captures the competition between potential assignments.

\[
P(O = o|v, w, \theta) = \begin{cases} 
P_O(v, w, \theta) & \text{if } o = \text{verb-theme-recipient} \\ 1 - P_O(v, w, \theta) & \text{if } o = \text{verb-recipient-theme} \end{cases}
\]  

(7)

where \(P_O(v, w, \theta)\) is the output of the syntactic alternant choice model described in Section 4.2.

\[
P(\Theta = \theta|v, w) = \begin{cases} 
P_\Theta(w) & \text{if } \theta = \{w_1 \mapsto \text{thm}, w_2 \mapsto \text{rec}\} \\ 1 - P_\Theta(w) & \text{if } \theta = \{w_1 \mapsto \text{rec}, w_2 \mapsto \text{thm}\} \end{cases}
\]  

(8)

where \(P_\Theta(w)\) is the output of the thematic role assignment model described in Section 4.3.

### 4.2 Syntactic alternant choice model

A key ingredient in the formulation of interpretability is the probability of a speaker choosing to use the dative alternant that she did. I conceive of the choice of alternant as the choice of an order for the thematic roles as realized by the NP arguments of the verb, so that the prepositional dative (NP-PP) is represented by the order \textit{verb-theme-recipient} and the double-object dative (NP-NP) is represented by the order \textit{verb-recipient-theme}. It follows from this treatment that I assume that the insertion of to in the prepositional dative case is triggered by the thematic role order \textit{verb-theme-recipient}; this does not reflect any strong theoretical commitment, but is rather a convenient assumption that clarifies the hypothesized connection between the phonetic realization of to and the degree to which the speaker’s intended interpretation can be accurately inferred by a listener.

Following previous studies (Bresnan et al., 2007; Bresnan and Hay, 2008; Bresnan and Ford, 2010), I constructed a model for the choice between alternants using binary mixed-effects logistic regression. This model attempts to predict the probability that the \textit{verb-theme-recipient} order of thematic roles will be chosen, given the verb sense, the assignment of thematic roles to arguments, and properties of those arguments, as outlined in Section 3.2.3. This probability is denoted by \(P_O(v, w, \theta)\)\textsuperscript{17}.

I hand-fit the model in a stepwise fashion, starting with a complete host of as many predictors as could reasonably be expected to be supported by the data without risk of overfitting (Harrell, 2001) and eliminating predictors one-by-one which did not meet statistical significance and did not contribute significantly to the explanatory power of the model, as measured by a likelihood ratio test\textsuperscript{18}. Following Bresnan et al. (2007), I included a random intercept

\textsuperscript{16}The formulation of interpretation probability allows for intrinsic properties of the verb to affect the thematic role assignment as well; however, as discussed in Section 4.3, the data obtained in this study was insufficient to reliably estimate such effects.

\textsuperscript{17}Properties of context and of the speaker were also investigated in the construction of this model and are discussed briefly in Appendix B; however, they were removed from the final model and thus do not appear in this notation for brevity.

\textsuperscript{18}In a likelihood ratio test, the goodness of fit of a complete model is compared to that of an equivalent model with a single predictor removed. This yields a measure of the degree to which the removed predictor adds explanatory value to the model, above and beyond that contributed by the other predictors, which can be tested for statistical significance.
for verb sense in all models because numerous verb senses occurred many times in the data, but did not include a random intercept for speaker identity because most speakers used very few datives. An investigation of the effects in different sub-corpora showed no reliable differences, indicating that the weights of constraints on alternant choice in the data collected here do not vary systematically as a function of time and/or situation (e.g. interview style and purpose).

Table 7 gives the coefficients of the fixed effects in the final model\textsuperscript{19}. A positive coefficient represents support for the \textit{verb-theme-recipient} order of thematic roles, while a negative coefficient represents support for the \textit{verb-recipient-theme} order. The intercept represents the bias towards the \textit{verb-theme-recipient} order of thematic roles when the theme is indefinite and nonpronominal, the recipient is animate and nonpronominal, and the theme and recipient have the same length. The effects are as expected based on previous work on the dative alternation (Bresnan et al., 2007; Bresnan and Hay, 2008; Bresnan and Ford, 2010); see Appendix B for further discussion.

The treatment of verb sense as a random effect means that the model generalizes beyond individual verb senses; nevertheless, the model does estimate effects of verb sense in a constrained form as additions to the intercept coefficient, drawn from a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 2.578$. These estimates are given in Table 8; as can be seen, some verb senses have quite strong biases in their thematic role order preferences. The verb sense effects seen here are consistent with those seen in Bresnan and Ford’s (2010) model of the dative alternation in American English; see Appendix B for details.

Investigation of the model showed that while it is affected by the large asymmetries in the data, it is still reliable. Under 10-fold cross-validation\textsuperscript{20}, the model assigned a higher probability to the order actually used than

19Note that 1605 observations were used to build the model, rather than the full 1613 in the corpus. The 8 observations left out had proper noun themes (\textit{show him Christchurch}), which appeared to pattern differently than definite common noun themes but were too few for their effect to be reliably modeled. Since it was judged inappropriate to fold proper nouns into the definite class, and since there were so few proper noun themes, constructions with proper noun themes were simply excluded from the data used to fit the model.

20In 10-fold cross-validation, the data is split into 10 groups. The model is then trained separately on each subset of 9 groups and used to make predictions on the remaining (unseen) group. The overall performance of the model is the average of its performances in each of these 10 independent testing situations.
Table 8: Random intercept additions for different verb senses.
Suffixes represent semantic class: transfer, future transfer, communication, prevention of possession, and abstract.

<table>
<thead>
<tr>
<th>Verb sense</th>
<th>Intercept</th>
<th>Verb sense</th>
<th>Intercept</th>
<th>Verb sense</th>
<th>Intercept</th>
<th>Verb sense</th>
<th>Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>tell.c</td>
<td>−4.850</td>
<td>read.c</td>
<td>−1.082</td>
<td>allow.a</td>
<td>−0.028</td>
<td>sell.t</td>
<td>1.175</td>
</tr>
<tr>
<td>teach.c</td>
<td>−3.525</td>
<td>offer.f</td>
<td>−0.804</td>
<td>award.f</td>
<td>−0.003</td>
<td>offer.a</td>
<td>1.454</td>
</tr>
<tr>
<td>charge.p</td>
<td>−3.277</td>
<td>show.c</td>
<td>−0.388</td>
<td>do.a</td>
<td>0.063</td>
<td>bring.a</td>
<td>1.837</td>
</tr>
<tr>
<td>lend.t</td>
<td>−2.107</td>
<td>serve.t</td>
<td>−0.349</td>
<td>hand.t</td>
<td>0.485</td>
<td>send.t</td>
<td>2.079</td>
</tr>
<tr>
<td>give.c</td>
<td>−2.032</td>
<td>give.t</td>
<td>−0.224</td>
<td>promise.a</td>
<td>0.739</td>
<td>allot.f</td>
<td>2.133</td>
</tr>
<tr>
<td>loan.t</td>
<td>−1.922</td>
<td>wish.a</td>
<td>−0.173</td>
<td>feed.t</td>
<td>0.823</td>
<td>take.a</td>
<td>2.786</td>
</tr>
<tr>
<td>cost.p</td>
<td>−1.833</td>
<td>pay.t</td>
<td>−0.155</td>
<td>send.e</td>
<td>0.859</td>
<td>write.t</td>
<td>2.836</td>
</tr>
<tr>
<td>give.a</td>
<td>−1.578</td>
<td>bring.t</td>
<td>−0.087</td>
<td>leave.a</td>
<td>0.886</td>
<td>pay.a</td>
<td>3.286</td>
</tr>
<tr>
<td>deny.p</td>
<td>−1.302</td>
<td>owe.t</td>
<td>−0.070</td>
<td>leave.f</td>
<td>0.912</td>
<td>make.e</td>
<td>3.363</td>
</tr>
<tr>
<td>owe.f</td>
<td>−1.110</td>
<td>swap.t</td>
<td>−0.048</td>
<td>lend.a</td>
<td>1.151</td>
<td>take.t</td>
<td>4.359</td>
</tr>
</tbody>
</table>

The probability of using the verb-theme-recipient order given a verb sense and arguments already assigned thematic roles, \( P_O(v, w, \theta) \), can be obtained from Tables 7 and 8 via (9):

\[
P_O(v, w, \theta) = \frac{1}{1 + \exp(-z_{v,w,\theta})} \tag{9}
\]

where

\( z_{v,w,\theta} = -0.723 \)

+ [random intercept addition for verb sense]

+ 1.634 × [relative complexity of recipient to theme]

+ [coefficients for all properties true of recipient / theme]

4.3 Thematic role assignment model

The second key ingredient in the formulation of interpretability is the probability of the argument NPs being assigned the thematic roles they are simply due to their intrinsic properties\(^{21}\). That is, assuming two arguments \( w_1 \) and \( w_2 \) were drawn at random for a given verb, with how much certainty could one be identified as a typical

\(^{21}\)The formulation of interpretation probability allows for the verb sense to impact thematic role assignment as well, but the data collected here were not sufficient to allow such effects to be reliably estimated.
recipient as opposed to a typical theme, or vice-versa? The model described in this section captures the probability that \( w_1 \) (the first argument randomly drawn in this manner, which may or may not appear first in the construction) is the theme and \( w_2 \) is the recipient, which I denote by \( P_\Theta(w) \). It does so by considering the properties that are typically true of recipients as opposed to themes, and vice-versa.

It is apparent from the data that animacy is an extremely strong factor in the assignment of thematic roles. Table 3 shows that animate entities are overwhelmingly likely to be recipients rather than themes; of the 1516 animate entities involved in dative constructions in the corpus, 1495 (98.6%) are recipients. Moreover, as Table 9 shows, when just one of the arguments is animate, that argument is the recipient 99.9% (1475/1476) of the time. The single case in the present data where the theme is animate and the recipient is inanimate, give her to something like a kindergarten, is clearly idiosyncratic: her refers to a papier-mâché pig and thus is not really animate. It therefore seems that, when the two arguments differ in animacy, the animate one is assigned the role of recipient near-categorically. This observation is supported by the fact that, of the 3078 datives with arguments of differing animacies in a database of syntactic alternations from spontaneous Canadian, British and American English created by Joan Bresnan and colleagues, just a single token has the animate entity as the theme (give her back to the Greyhound Association, where her refers to a greyhound\(^{22}\)). Given this observation, I decided to have the thematic role assignment model first filter the data based on animacy, assigning a \( P_\Theta(w) \) value of 0 in cases where \( w_1 \) is animate and \( w_2 \) is inanimate and a \( P_\Theta(w) \) value of 1 in cases where \( w_1 \) is inanimate and \( w_2 \) is animate.

In the subset of the data where the arguments have the same animacy (137 tokens), it was found that definite proper nouns were always assigned the role of recipient (7 tokens). Given that all instances of definite proper nouns as themes occurred in cases where the arguments had opposing animacies (e.g. show him Christchurch), I decided to have the thematic role assignment model further filter the data and assign a \( P_\Theta(w) \) value of 0 in cases where the arguments agree in animacy and \( w_1 \) is a definite proper noun and a \( P_\Theta(w) \) value of 1 in cases where the arguments agree in animacy and \( w_2 \) is a definite proper noun.

For the remaining 130 tokens not assigned \( P_\Theta(w) \) values by this filtering process, I considered the associations between other properties of the arguments and their thematic roles. The association is strongest for definiteness; as shown in Table 10, in 72 of the 78 cases where the two arguments differ in definiteness (92.3%), the definite one is the recipient. While there are associations between short, pronominal and plural arguments and the role recipient,

<table>
<thead>
<tr>
<th></th>
<th>Recipient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Animate</td>
</tr>
<tr>
<td>Theme</td>
<td>Animate</td>
</tr>
<tr>
<td></td>
<td>Inanimate</td>
</tr>
</tbody>
</table>

\(^{22}\) Even this example is questionable, as both pets and organizations have intermediate levels of animacy (Garretson, 2004).
Table 10: Definiteness of theme and recipient together in animacy- and proper noun-filtered data.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Recipient</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Indefinite</td>
<td>Definite</td>
</tr>
<tr>
<td>Indefinite</td>
<td>8</td>
<td>72</td>
</tr>
<tr>
<td>Definite</td>
<td>6</td>
<td>44</td>
</tr>
</tbody>
</table>

in line with observations from Figure 1 and Tables 2 and 5, these associations are likely highly overlapping (as pronouns are short, and the vast majority of plural recipients consist of the pronominal them) and thus require a substantial amount of data to be disentangled. They also overlap highly with the definiteness effect (as pronouns are definite), to the point that asymmetries in thematic roles based on length, pronominality and number fail to exist in the subset of the filtered data where the arguments have the same definiteness. Because of these observations, and because of the risk of overfitting due to the small number of cases with indefinite recipients and definite themes, I decided to model the only effect of definiteness on thematic role assignment in this subset.23

To model the effect of definiteness on thematic role assignment (after accounting for the effects of animacy and proper nouns) in a way that would extend beyond the available sample of 130 tokens, I pursued a bootstrap aggregation approach. I drew 1000 bootstrap samples of 130 tokens from the filtered data, with replacement (so that the same token may be included in a given bootstrap sample multiple times or not at all). Within each bootstrap sample, I randomly selected 65 tokens in which I indexed the recipient as \( w_1 \) and the theme as \( w_2 \), and applied the reverse indexing in the other 65 tokens. I then created a variable measuring the definiteness of \( w_2 \) relative to \( w_1 \), defined in (10). For each bootstrap sample, I ran a logistic regression predicting the probability of the thematic role assignment \( \{ w_1 \mapsto \text{thm}, w_2 \mapsto \text{rec} \} \) from the relative definiteness of the arguments.24 Finally, I averaged the coefficient estimates across all 1000 bootstrap samples to arrive at a final model, shown in Table 11.25

Relative definiteness =

\[
\begin{cases} 
-1 & \text{if } w_1 \text{ is definite and } w_2 \text{ is indefinite} \\
0 & \text{if } w_1 \text{ and } w_2 \text{ agree in definiteness} \\
1 & \text{if } w_1 \text{ is indefinite and } w_2 \text{ is definite}
\end{cases}
\]

23Focusing solely on definiteness no doubt misses a lot of detail, yielding a very coarse thematic role assignment model. For example, it lacks access to semantic and pragmatic content, which are intuitively important in determining the plausibility of thematic role assignments (at least in offline judgments). While this may make the model inadequate for use in the absolute evaluation of a construction’s interpretability, it is still useful for indicating the relative interpretability of different constructions.

24I also ran a version of the model where the effect of relative definiteness was conditioned on verb sense via a random slope; however, the verb sense effects were unduly influenced by a small number of non-representative cases with organizations treated as animate entities. Because these effects thus reflected animacy and not definiteness, I concluded that there were too few tokens in the filtered subset to reliably estimate differential effects of relative definiteness on thematic role assignment for different verb senses.

25Conservative two-sided \( p \)-values for a bootstrapped statistic \( x \) such as the coefficients in Table 11 can be estimated by \((2M + 1)/(N + 1)\), where \( M = \min(\# \text{ of bootstrap samples in which } x \leq 0, \# \text{ of bootstrap samples in which } x \geq 0)\) and \( N \) is the total number of bootstrap samples.
Table 11: Bootstrap aggregated logistic regression model predicting the probability that the \( \{w_1 \mapsto \text{thm}, w_2 \mapsto \text{rec}\} \) assignment of thematic roles is used.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Min. bs value</th>
<th>Max. bs value</th>
<th>Std. Deviation</th>
<th>p-bs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.007</td>
<td>-0.558</td>
<td>0.523</td>
<td>0.164</td>
<td>0.9600</td>
</tr>
<tr>
<td>Relative definiteness of ( w_2 ) to ( w_1 )</td>
<td>2.631</td>
<td>1.387</td>
<td>20.585</td>
<td>0.907</td>
<td>0.0010 ***</td>
</tr>
</tbody>
</table>

Num. obs = 130; Num. bootstrap samples = 1000

\[p^* < 0.001, \; \cdot \; p^* < 0.01, \; * \; p^* < 0.05, \; . \; p^* < 0.1\]

Probabilities can be obtained from the model in a way analogous to that described by (9) for the syntactic alternant choice model. In the filtered subset of the data, when \( w_1 \) is definite and \( w_2 \) is indefinite, \( w_1 \) is assigned the role theme with probability \( P_{\Theta}(w) = 0.07 \), and when \( w_1 \) is indefinite and \( w_2 \) is definite, \( w_1 \) is assigned the role theme with probability \( P_{\Theta}(w) = 0.93 \). Unsurprisingly, when \( w_1 \) and \( w_2 \) agree in definiteness, \( P_{\Theta}(w) = 0.50 \) and the assignment of thematic roles is completely at chance. Figure 2 shows a decision tree specifying the value that should be assigned to \( P_{\Theta}(w) \) in different conditions.

Figure 2: Decision tree outlining the thematic role assignment model probabilities.

\[
\begin{align*}
\text{Argument animacies} & \\
\text{Which argument is animate?} & \\
\text{different} & \\
\text{same} & \\
\text{Is there a proper noun?} & \\
\text{no} & \\
\text{yes} & \\
\text{Argument definitenesses} & \\
\text{different} & \\
\text{same} & \\
\text{Which argument is definite?} & \\
\text{Which argument is a proper noun?} & \\
\text{\( P_{\Theta}(w) = 0.50 \)} & \\
\text{\( P_{\Theta}(w) = 0 \)} & \\
\text{\( P_{\Theta}(w) = 1 \)} & \\
\end{align*}
\]

\( P_{\Theta}(w) = 0.07 \) \( P_{\Theta}(w) = 0.93 \)

4.4 Assessing the impact of interpretability on phonetic cues

In this investigation, I am concerned with the degree to which the speaker will strengthen phonetic cues to the word \( to \) in a prepositional dative if she thinks that the listener may not correctly infer her intended assignment of thematic roles to NP arguments in the dative on the basis of the properties of the arguments alone. In cases
where the listener is categorically able to correctly infer the intended thematic roles solely from properties of the arguments, however, interpretability need not be attended by the speaker and phonetic cues to *to* are thus expected to vary widely for other reasons (such as context and natural speech patterns). To avoid the potential for such variation to mask the effect of interpretability, I excluded all such cases from consideration; this means that I am left to consider how interpretability affects phonetic cues to *to* in the prepositional datives among the 130 cases identified as having uncertain thematic role assignments in Section 4.3.

This set of 130 utterances contains 61 prepositional datives. I excluded twelve of these utterances because the *to* was bounded by a pause, as in such cases the word duration and segmental realizations may be misrepresentative (Fox Tree and Clark, 1997; Shriberg, 2002). I did not exclude utterances where there was a disfluency elsewhere in the construction; in particular, I included 9 utterances where there was a disfluency (in most cases a pause, but in some cases a filler or repeated word) after the determiner following *to*, where the realization of *to* is likely to be unaffected by the disfluency because of the intervening word. The exclusion procedure left 49 utterances in which the effect of interpretability on phonetic cues to *to* can be investigated. This investigation takes the form of linear regression, where interpretability is used to predict the strength of a phonetic cue.

The phonetic cue that I consider here is duration; that is, I investigate the hypothesis that the length of the *to* and of its constituent segments in a prepositional dative will decrease as interpretability increases. Duration is just one of many possible phonetic cues which could see an effect of interpretability; it is used primarily because it is simple, one-dimensional, and has been used widely in corpus and experimental studies to capture effects of informativity and processing within the framework of a probability-sensitive grammar (Aylett and Turk, 2004, 2006; Bell et al., 2009; Gahl and Garnsey, 2004; Gahl et al., 2006; Jurafsky et al., 2002; Kuperman and Bresnan, 2012; Tily et al., 2009).

Of course, duration is open to influences other than interpretability. Both words and segments are known to have different durations depending on their context: a given word is shorter when it is predictable from its surrounding words (Bell et al., 2009; Jurafsky et al., 2002) and segments have different durations depending on their surrounding segments (House, 1961; Umeda, 1977). Additionally, words and segments in speech produced at a fast rate are (by definition) shorter than those in speech produced at a slow rate. It is important to control for factors such as these in order to be sure that any observed effect of interpretability exists above and beyond other expected influences on duration.

I controlled for speech rate by multiplying the duration of interest by the local speech rate calculated in (3). This yielded a measure reflecting how many syllables the duration represents, which is independent of the rate at which those syllables may be spoken.

I controlled for predictability of *to* from surrounding words by including in the regression analysis the logs of the conditional probabilities of *to* given the word before it and the word after it, calculated in (4) and (5) respectively.

---

26 Note, however, that some of these utterances are not able to be investigated for all phonetic cues due to poor quality.
Including these controls in the regression ensures that the effects of interpretability are above and beyond those of predictability. Note, however, that it decreases the power of the model as a whole; due to this and the small number of datapoints, some F-tests for overall models indicated that all predictors taken together do not have a statistically significant effect on the response, even while some of the predictors showed significant individual effects based on the t-test of their coefficients. It is assumed that such a situation simply reflects limitations of the small dataset, so I tentatively base my conclusions on the predictor-wise t-tests rather than the overall F-tests.

Controlling for segmental environment was only possible for the segment immediately preceding to; the segments immediately following to were dominated by /ð/ due to 30 of the 49 recipients beginning with the, making it impossible to robustly estimate effects for any meaningful cluster of following segments due to data sparsity. I considered two ways of controlling for the previous segment, both of which took the form of adding to the regression model a predictor that partitioned the data based on the class of the segment immediately preceding to. Additionally, I included an interaction between the class of the preceding segment and interpretability, under the assumption that speakers may respond to interpretability pressures differently in different phonological environments. Such responses are expected to be localized to the /t/ in to, as this is the part of to that has its environment provided by the preceding segment. The first partition of the classes into segments is based on similarity in place of articulation and separates the segments into alveolars (/t, d, s, z, n/) and non-alveolars. The idea behind this partition is that a speaker may choose to lengthen the /t/ in to in low interpretability when doing so is convenient because it requires no extra articulatory effort, since the tongue is already in position for producing /t/. The second partition of the segments into classes is based on perceptual similarities and separates the segments into those that are perceptually confusable with /t/ by virtue of differing in up to one (coarse) feature (/p, t, k, d, s/) and those that are distinct from /t/ and differ in multiple features. Bailey and Hahn (2005) argue that this partition is the most effective way of capturing phoneme confusability, and the posited set of confusable consonants are readily confusable with /t/ in CV syllables in noise (Miller and Nicely, 1995).

As durations and probabilities are often subjected to a negatively-accelerating increasing transformation (e.g. log-transform) in linear regressions in order to control high-leverage points (i.e. suppress excessive influence of extreme values which may be driving effects) and bring the distribution of residuals closer to the normality assumed by the model\textsuperscript{27}, I considered such transformations in my analysis. For each of duration and interpretability, in addition to the untransformed data, I considered the square-root-transformed data $x' = \sqrt{x}$ and the log-transformed data $x' = \log(x)$, yielding a total of 9 possible pairs of data vectors. I created a linear regression model using each pair of data vectors (and controls already mentioned) and inspected the distribution of residuals and the fit of

\textsuperscript{27}It is commonly stated in the psycholinguistic literature that such a transformation is conducted in order to bring the distribution of the data closer to normality; however, this is unnecessary, as regression models can fit to data of arbitrary distributions – consider, for example, the simple problem of predicting a variable distributed following $y = Z + \epsilon$ (where $\epsilon \sim N(0, 0.01)$) from a predictor distributed following $x = Z$, where $Z$ is uniformly distributed on $[0, 1]$. What the regression model assumes is that the residuals, or errors of the prediction, are normally distributed, and while normality of the residuals may in practice be easier to obtain with normally-distributed data, it can also be obtained with non-normally-distributed data. The key consideration in transformations of the data (at least from a statistical standpoint) is that the transformation is appropriate for the relationship displayed between the response and the predictors in the data; that is, that the transformed variables are approximately linearly related to each other.
the model as measured by $R^2$. I chose a pair of transformations which yielded a distribution of residuals close to normality and a good model fit, giving preference to pairs where both variables were transformed for reasons of leverage control. In all cases, the best transformation for duration was the square-root transformation and the best transformation for interpretability was the log transformation.

To ensure that the variable transformations effectively controlled for high-leverage points which could be artificially driving effects, I calculated Cook’s distance $D_i$ for each datapoint in the model. In all cases, $D_i < 0.7$, indicating no extreme influence of a single datapoint (Cook and Weisberg, 1982).

To summarize, I assessed the impact on interpretability on phonetic cues by regressing rate-normalized square-root duration against log interpretability, the class of the segment preceding to, their interaction, and the predictability of to from surround words. The transformations of duration and interpretability were chosen to allow for accurate and appropriate modeling. All numerical predictors in the regression were centered to aid model fitting.

For visualization purposes, I converted the output of the model into the natural millisecond units and transformed the original datapoints to the same scale (normalizing for speech rate and predictability of to from context). A description of this process can be found in Appendix C.

5 Results

5.1 to-duration and interpretability

Table 12 gives the results of the regression of to-duration, specifying phonological environment as whether the immediately preceding segment is perceptually confusable with /t/ (/p, t, k, d, s/). As can be seen from the report for log(Interpretability), interpretability has no significant effect on the duration of to when the preceding segment is confusable with /t/, after controlling for predictability and phonological environment. There is no significant main effect of Preceding segment, indicating that, on average and after controlling for predictability and interpretability, instances of to in different phonological environments do not have different lengths. The lack of a significant interaction between log(Interpretability) and Preceding segment shows that to-duration does not pattern differently with interpretability based on phonological environment; this implies that interpretability has no effect on the duration of to when the preceding segment is not confusable with /t/, which is confirmed statistically ($t = 0.577, p = 0.5669$; not shown in the table). These results are illustrated in Figure 3.

Equivalent results were obtained when specifying phonological environment as whether the immediately preceding segment is alveolar. This indicates that, no matter how the controls are conceptualized, interpretability does not affect the duration of to as a whole.

Of course, the fact that to does not significantly lengthen with interpretability does not mean that interpretability has no effect on the strength of phonetic cues in to (as summarized by duration). If the two phones in to have difference acoustic saliences, speakers could achieve phonetic strengthening of to simply by extending one of them.
Table 12: Linear regression model predicting square-root rate-normalized to-duration. The intercept captures duration when the variables for interpretability and predictability of to take on their average values over the dataset and the preceding word ends in a segment that is confusable with /t/.

|                                | Estimate | Std. Error | t-value | Pr(>|t|) |
|--------------------------------|----------|------------|---------|----------|
| (Intercept)                    | 0.842    | 0.0339     | 24.852  | < 2 × 10^{-16} *** |
| log(Interpretability)          | -0.100   | 0.061      | -1.645  | 0.1073   |
| Preceding segment = non-confusable | -0.037   | 0.050      | -0.732  | 0.4680   |
| log(to-predictability from previous word) | 0.029    | 0.023      | 1.278   | 0.2082   |
| log(to-predictability from following word) | -0.069   | 0.034      | -2.030  | 0.0485 * |
| log(Interpretability) × Preceding segment = non-confusable | 0.153    | 0.110      | 1.39    | 0.1716   |

Multiple R^2 = 0.197; F(5, 43) = 2.11; p = 0.0826; Num. obs. = 49

\*\*\* p < 0.001, \*\* p < 0.01, \* p < 0.05, \· p < 0.1

Figure 3: Partial effect plots of the predicted effect of interpretability on to-duration: (A) after segments which are perceptually confusable with /t/ (p, t, k, d, s); (B) after segments which are not perceptually confusable with /t/. Speech rate and predictability controls have been normalized to their average values over the dataset. Circles show datapoints after normalization, the solid red line shows the partial effect of interpretability, and dashed red lines show 95% confidence intervals. Filled circles represent utterances with a disfluency after the determiner following to. Labels show the word preceding to.
This could happen in two ways without significantly altering the length of to as a whole: one phone could be lengthened while the other is shortened, or the lengthening of one phone could be substantial relative to the average length of that phone but small relative to the duration of the entire word. For this reason, it is worthwhile to consider a relationship between interpretability and the /t/ or the vowel in to.

5.2 Vowel duration and interpretability

Table 13 and Figure 4 show the results of a regression of vowel duration in to. It is quite clear that there are no differences in the duration of the vowel based on interpretability or even on any controls. Specifying phonological environment based on preceding alveolar segments rather than confusable ones did not cause any differences to appear (which is not surprising, as the segment preceding to does not provide an environment for the vowel); the duration of the vowel in to is simply unpredictable in this dataset.

However, comparing Figure 4A to Figure 3A reveals an interesting observation: instances of to in low-interpretability utterances appear to have relatively shorter vowels (in comparison to those in mid- and high-interpretability utterances) than would be expected based on their overall duration. This indicates that, though the duration of the vowel doesn’t change with interpretability, the duration of /t/ might.

![Figure 4](image_url)

Figure 4: Partial effect plots of the predicted effect of interpretability on vowel duration in to: (A) after segments which are perceptually confusable with /t/ (p, t, k, d, s); (B) after segments which are not perceptually confusable with /t/. Speech rate and predictability controls have been normalized to their average values over the dataset. Circles show datapoints after normalization, the solid red line shows the partial effect of interpretability, and dashed red lines show 95% confidence intervals. Filled circles represent utterances with a disfluency after the determiner following to. Labels show the word preceding to.
Table 13: Linear regression model predicting square-root rate-normalized vowel-duration in to. The intercept captures duration when the variables for interpretability and predictability of to take on their average values over the dataset and the preceding word ends in a segment that is confusable with /t/.

|                      | Estimate | Std. Error | t-value | Pr(>|t|) |
|----------------------|----------|------------|---------|----------|
| (Intercept)          | 0.603    | 0.0285     | 21.170  | < 2 × 10^{-16} *** |
| log(Interpretability)| −0.002   | 0.045      | −0.048  | 0.9617   |
| Preceding segment = non-confusable | −0.068   | 0.041      | −1.665  | 0.1049   |
| log(to-predictability from previous word) | −0.010   | 0.017      | −0.588  | 0.5603   |
| log(to-predictability from following word) | −0.016   | 0.026      | −0.602  | 0.5513   |
| log(Interpretability) × Preceding segment = non-confusable | 0.035    | 0.084      | 0.420   | 0.6772   |

Multiple $R^2 = 0.1538$; $F(5, 35) = 1.272$; $p = 0.2979$; Num. obs. = 42

5.3 /t/-duration and interpretability

The suspicions arising in Section 5.2 prove to be correct: as Table 14 and Figure 5 show, there is a significant effect of interpretability on /t/-duration, at least when to is preceded by a segment that is perceptually confusable with /t/. As the probability of a listener correctly interpreting the meaning the speaker intends to communicate in a prepositional dative construction decreases, the speaker produces the /t/ in to substantially longer, as shown by the fact that the coefficient for log(Interpretability) is negative and reasonably large relative to the intercept (which represents baseline or average /t/-duration). The increase in /t/-duration between high- and low-interpretability utterances is approximately 60ms, which is undoubtedly perceptible. While the effect of interpretability changes significantly based on the phonological environment, as shown by the interaction between log(Interpretability) and Preceding segment, it is not the case that a reliable relationship of the opposite type is seen when to is preceded by a segment that is not confusable with /t/; rather, the effect of interpretability is not significant in this case ($t = 1.05, p = 0.3000$; not shown in the table). Qualitatively identical results are seen when the phonological environment is specified based on the place of articulation of the segment preceding to (not shown).
Table 14: Linear regression model predicting square-root rate-normalized /t/-duration, with phonological environment specified as whether the preceding segment is perceptually confusable with /t/ (/p, t, k, d, s/). The intercept captures duration when the variables for interpretability and predictability of to take on their average values over the dataset and the preceding word ends in a segment that is confusable with /t/.

|                          | Estimate | Std. Error | t-value | Pr(>|t|) |
|--------------------------|----------|------------|---------|----------|
| (Intercept)              | 0.639    | 0.033      | 19.299  | < 2 × 10^{-16} *** |
| log(Interpretability)    | -0.127   | 0.058      | -2.234  | 0.0313 * |
| Preceding segment = non-confusable | -0.014    | 0.050    | -0.288  | 0.7748   |
| log(to-predictability from previous word) | 0.028    | 0.024 | 1.160 | 0.2530 |
| log(to-predictability from following word) | -0.055 | 0.033 | -1.678 | 0.1013 |
| log(Interpretability) × Preceding segment = non-confusable | 0.226 | 0.108 | 2.089 | 0.0433 * |

Multiple $R^2 = 0.223; F(5, 39) = 2.23; p = 0.0702; \text{Num. obs.} = 45$

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$

Figure 5: Partial effect plots of the predicted effect of interpretability on /t/-duration: (A) after segments which are perceptually confusable with /t/ (p, t, k, d, s); (B) after segments which are not perceptually confusable with /t/. Speech rate and predictability controls have been normalized to their average values over the dataset. Circles show datapoints after normalization, the solid red line shows the partial effect of interpretability, and dashed red lines show 95% confidence intervals. Filled represent utterances with a disfluency after the determiner following to. Labels show the word preceding to.
5.4 Non-acoustic reflexes of interpretability

I conclude the results section by presenting a non-acoustic result from the investigation, discovered in the preparation for the regressions. This result concerns the relationship between interpretability and the tendency to use the prepositional dative (as opposed to the double-object dative).

As noted in Section 4.4, prepositional datives constitute nearly half (61 out of 130) of the datives with uncertain interpretations (where assignment of thematic roles cannot be predicted with certainty based merely on the properties of the arguments, as both are common nouns with the same animacy). Given that there are seven times as many double-object datives as there are prepositional datives in the corpus, this means that prepositional datives are overrepresented among datives with uncertain interpretations ($\chi^2 = 48.91, p = 2.679 \times 10^{-12}$). That is, a speaker is more likely to encode a $(verb, recipient, theme)$ triple in a prepositional dative (compared to a double-object dative) when the theme could be interpreted as a recipient or vice-versa than when the assignment of theme and recipient roles in interpretation is clear. This may be because the use of $to$ in prepositional datives gives an extra cue to the intended interpretation and thus allows it to be more robustly transmitted.

An alternative interpretation of this result is that the similarity of the two arguments (due to the fact that they have the same animacy) leads them to interfere with each other in memory (Gennari et al., 2012; MacDonald, 2013), so the producer chooses the prepositional dative in order to separate them more in the production plan and thus reduce this interference. Given that Kuperman and Bresnan (2012) present results suggesting that the entire dative construction is planned at the point of articulation of the verb, this account is feasible. To the extent that formal definiteness also bears on argument similarity, this account additionally predicts that arguments with the same animacy and definiteness should create even more interference, motivating an even greater proportion of prepositional datives in the corresponding subset of the data. As Figure 6 shows, this is seen in the data: 37 of the 52 datives with arguments with the same animacy and definiteness are prepositional, which is a significantly greater proportion than in all datives with arguments with the same animacy ($\chi^2 = 7.83, p = 0.0051$).

Though there seems to be support for the interference-reduction account of the proportion data, it does not invalidate the support for the interpretability account because the constructions realized as prepositional datives tend to have lower interpretability than those realized as double-object datives. Figure 7 demonstrates this for the subset of datives where the arguments have the same animacy and definiteness, where interference reduction predicts a greater chance of using the prepositional alternant. It is clear that the favor towards the prepositional alternant is not equal across the interpretability range; rather, it grows as interpretability lowers ($t = -6.76, p = 1.43 \times 10^{-8}$). Thus, an effect of interpretability exists above and beyond any effect of interference reduction.
Figure 6: Proportions of datives that have the prepositional form in different subsets of the data, specified by constraints on the arguments. All constructions subject to the stated constraints were included for this plot (not just those with uncertain interpretations, as in the text). Whiskers show 95% confidence intervals.

Figure 7: Histograms comparing the distribution of interpretability values for double-object (NP-NP) and prepositional (NP-PP) datives where the arguments have the same animacy and definiteness.
6 Results revisited

6.1 Duration and planning

On inspecting Figure 5, an intriguing observation arises: the two datapoints with lowest interpretability had disfluencies after the determiner following to. Under the standard account of disfluencies (see review in Rochester, 1973), this indicates that the speaker experienced difficulty in planning the retrieval or articulation of the second argument in those datives. Could this difficulty cause the lengthening of /t/?

It certainly seems that planning effects could be relevant for segmental durations in to: as can be observed from Figure 4, when there is a disfluency after the determiner following to, the vowel in to is longer (independent of interpretability; \( t = -2.460, p = 0.0187 \)). However, no such wholesale effect is seen for /t/-duration across all disfluent instances (\( t = -0.831, p = 0.4106 \)); rather, it is only in the lowest-interpretability datives that the effect is seen. As Figure 8 shows, /t/-duration grows as interpretability lowers even among instances of to in datives which show disfluencies. This indicates that the planning difficulty account is insufficient, as it cannot account for the interaction of interpretability with /t/-duration.

Of course, not all planning difficulties are equal, so this dismissal of the planning difficulty account may be premature. It could be the case that the low-interpretability datives examined here induced greater planning difficulties, which caused /t/-lengthening in addition to vowel-lengthening. There are two ways in which this could have happened: directly, where the properties that led these utterances to have low interpretability also led them to be difficult to plan; or indirectly, where properties other than those related to interpretability caused extreme planning difficulty in these utterances, and the fact that they also happened to have low interpretability is merely due to chance.

Given that low-interpretability datives have low probabilities of syntactic alternant choice and/or thematic role assignment (by definition), and that such high-level probabilities have been shown to affect the use of disfluencies (Tily et al., 2009), the direct account is plausible. It is also supported by the fact that, in the data examined here, the proportion of datives with disfluencies following the determiner after to grows as interpretability decreases. In high-interpretability datives (interpretability \( \geq 0.75 \)), 1 of 15 (6.7%) utterances exhibited disfluencies; in mid-interpretability datives (interpretability between 0.25 and 0.75), the corresponding proportion was 6 of 28 (21.4%); and in low-interpretability datives (interpretability \( \leq 0.25 \)), disfluencies were seen in both (100%) of the utterances. However, it is unclear exactly how low-interpretability datives should give rise to planning difficulty that is concentrated on the second argument.

The indirect account is plausible but not supported by the data. Two proposals for the indirect account are that greater planning difficulties could be due to interference from similarity between arguments (Gennari et al., 2012; MacDonald, 2013) or due to the complexity of the second argument (Clark and Wasow, 1998). However, neither of these proposals hold of the two lowest-interpretability datives (selling it to a Brisbane company, selling...
Figure 8: Partial effect plots of the predicted effect of interpretability on /t/-duration in to: (A) when there is a disfluency after the determiner following to; (B) when the dative is fluent. Speech rate and predictability controls have been normalized to their average values over the dataset. Circles show datapoints after normalization, the solid red line shows the partial effect of interpretability, and dashed red lines show 95% confidence intervals. Labels show the word preceding to.

6.2 Possible data limitations

If interpretability is posited not to relate to planning difficulty, then the fact that the interpretability effect can be captured solely across the instances where a disfluency followed the determiner after to (as seen in Figure 8) suggests that limitations of the data could be obscuring conclusions that can be drawn. There are multiple ways of specifying the conditions under which the effect exists: after segments that are perceptually confusable with /t/; after alveolar segments; in cases where the speaker experiences planning difficulties; or even, as can be inferred from Figure 5, simply after the word it. It is hard to distinguish between these because the partitions they create in the data are highly overlapping: for example, the final segment of it is both alveolar and perceptually confusable with
/t/, and many of the datives with it as theme have disfluencies after the determiner following to. Furthermore, as each of these partitions include the two datapoints with lowest interpretability (which both have long /t/s), it may simply be that the effect is due to these two points. To test the hypothesis that the two datapoints corresponding to the lowest-interpretability datives are single-handedly causing the effect, I created 10,000 random partitions of the data into two groups, one of which (the positive group) always contained these two datapoints. The number of additional members of the positive group in each partition was drawn from a normal distribution over the total number of additional datapoints. For each partition of the data generated in this way, I ran a regression model analogous to those previously reported and stored the t-statistic corresponding to the interpretability effect in the positive group. 3055 (30.55%) of these partitions generated a t-statistic less than −1.96, corresponding to p-values less than 0.05 (if no correction is made for multiple testing) for negative effects of interpretability on /t/-duration. The fact that most of the partitions did not yield statistically significant effects of interpretability indicates that the two lowest-interpretability datapoints are not systematically driving the observations on their own; however, the fact that nearly one third of the partitions involving these two datapoints did yield statistically significant effects of interpretability suggests that they are highly influential. Unfortunately, it is not tenable to remove them from the analysis because they are the only observations of low-interpretability datives in the corpus. More observations would be required to determine whether they are representative of low-interpretability datives and whether the effects of interpretability are real or just an artifact of the data at hand.

6.3 Comparison of interpretability to its components

Recall that interpretability is a function of the probabilities of syntactic alternant choice and thematic role assignment. To be sure that the effects found here are specific to interpretability and not instead due to either of its component probabilities, I ran analogous regressions to those in Section 5.3, using the (log-transformed) component probabilities in the place of interpretability.

The regression using syntactic alternant choice probability showed no significant effect of this probability on /t/-duration (t = −0.231, p = 0.8187 after segments confusable with /t/; other data partitions equivalent). This is not surprising, as the most influential datapoints in yielding the interpretability effect were datives with the pronoun it as the theme, which received near-certain values from the syntactic alternant choice model (see Table 7). As a result, most of the variation in /t/-duration is concentrated at extremely high syntactic alternant choice probabilities. The absence of an effect of syntactic alternant choice probability on /t/-duration is consistent with the absence of a likewise effect on to duration reported by Kuperman and Bresnan (2012) for prepositional datives in American English and indicates that the observed effect of interpretability is not due to the predictability of the signal from the point of view of the speaker.

28While the Cook’s distance analysis in Section 4.4 indicated that no individual datapoint had extreme influence on any regression results, it is entirely possible that this pair of datapoints did.
However, there was a significant effect of thematic role assignment probability on /t/-duration ($t = -2.191, p = 0.0345$ after segments confusable with /t/; other data partitions equivalent). That is, the /t/ in to was longer (in certain environments) when the speaker assigned thematic roles to arguments in a way that was at odds with the general biases observable from other assignments to arguments with the same properties. This is also not surprising, as thematic role assignment probability forms the prior for interpretability (see equation (6)) and is highly correlated with it ($r = 0.748$ over the subset of data included in the /t/-duration model). It is unclear how the existence of this effect should be interpreted, as thematic role assignments are a property of the message, not the signal. Low-probability thematic role assignments reflect unusual messages, which could cause signal-planning difficulty for the speaker due to their unfamiliarity (see Section 6.1). But comprehending unusual messages is also difficult for the listener, so the effect of thematic role assignment probability may instead be a crude reflection of audience design considerations. While interpretability yielded a slightly stronger effect and a better model fit than thematic role assignment probability ($t = -2.234$, multiple $R^2 = 0.223$ and AIC = $-42.46$ for interpretability vs. $t = -2.191$, multiple $R^2 = 0.199$ and AIC = $-41.10$ for thematic role assignment probability), the difference is very small because of the high degree of correlation between the two metrics, and data sparsity prevents interpreting the difference as reliably indicating the superiority of the interpretability metric. Thus, the statistics cannot determine whether the observed effect of interpretability really is due to audience design, or whether it is instead due to listener-insensitive planning difficulty.

7 General Discussion

To briefly recap, regression analyses of corpus data showed that the /t/ in the critical word to in prepositional datives lengthens as the interpretability of the dative decreases, provided the word before to ends in a segment which is perceptually confusable with /t/ or shares its place of articulation (both environments separately yield the effect). There was no indication of any such effect on the vowel in to, and while the duration of to itself showed the same pattern, it was not statistically significant. Additionally, the prepositional dative alternant was seen to be favored more as interpretability lowered.

These results are consistent with the hypothesis that speakers modulate cues in the linguistic signal to avoid interpretation ambiguities, as predicted by an information-based treatment of communication with consideration of audience design. Encoding a message in a prepositional dative alternant allows the critical word to to provide an additional pointer to the speaker’s intended interpretation, which is valuable when the probability of the listener otherwise inferring that interpretation (without to) is low. Though to is completely predictable and thus uninformative for the speaker once she has committed to the prepositional alternant, it is highly informative to the listener in low-interpretability cases. Through experience of being a listener in other situations, the speaker is aware of this informativity and thus exhibits audience design, encoding a low-interpretability message in the prepositional
alternant and strengthening phonetic cues to *to* in order to increase the chance of communicative success. That this strengthening takes the form of lengthening the */t/* but not the vowel is consistent with a related awareness that the initial phone of a word is highly informative of the word’s identity (van Son and Pols, 2003). The context-sensitivity in the effect is also consistent with an information-based treatment: lengthening */t/* after a segment with which it is perceptually confusable increases robustness more than lengthening it after a non-confusable segment (through ensuring that the cues to */t/* do not get misattributed to the previous segment), yielding a greater motivation for expending the effort required to exhibit audience design; and lengthening */t/* after an alveolar segment requires less effort than lengthening it after a non-alveolar segment (because it does not require moving the tongue), yielding a decreased barrier to exhibiting audience design.

A separate regression using syntactic alternant choice probability as the primary predictor indicated that predictability from the point of view of the speaker cannot account for variance in */t/*-duration, in agreement with Kuperman and Bresnan’s (2012) findings for *to*-duration in prepositional datives in American English. However, a similar regression showed that the other component of interpretability, thematic role assignment probability, *can* account for some of this variance. While this may be a crude reflection of audience design, it could also be a symptom of the speaker’s planning difficulty. Low-probability thematic role assignments give rise to low interpretability, and it may be posited that they also induce planning difficulty due to their unfamiliarity. Thus, the lengthening of */t/* as interpretability decreases may reflect the difficulty of planning a signal with an unusual underlying message. This claim is supported by the fact that disfluencies become increasingly common as interpretability decreases.

However, there are several aspects of the results that cannot be satisfactorily explained by the claim that interpretability is a proxy for listener-insensitive planning difficulty. Firstly, the locus of the interpretability-based lengthening on the */t/* but not the vowel of *to* is only explained by planning difficulty if this difficulty is resolved between the two phones. However, the vowel in *to* is significantly longer when the realization of the argument following *to* contains a disfluency, indicating that clear cases of planning difficulty exhibit effects throughout *to*. Secondly, in these cases with disfluencies that clearly exhibit planning difficulty, the */t/* in *to* is *not* systematically lengthened like the vowel; rather, it is only lengthened when interpretability is low. When interpretability is high and there is a disfluency in the second argument, */t/* is, in fact, relatively short (compared to in high-interpretability cases without disfluencies). If interpretability is simply a proxy for planning difficulty, then we would expect */t/-duration to be similar in all of these cases of clear difficulty. Thirdly, the claim that interpretability merely reflects the difficulty of planning a signal for an unusual message fails to explain why the prepositional dative is favored more as interpretability decreases, as it is unclear how a prepositional dative should ease the burden of such planning more than a double-object dative. Thus, though there seems to be a relationship between interpretability and planning difficulty, the claim that interpretability is a proxy for the difficulty of planning a signal for an unusual message is incompatible with the results.

The claim that prepositional alternant selection and *to*-lengthening are driven by the avoidance of interference
between similar arguments in memory (Gennari et al., 2012; MacDonald, 2013) is also incompatible with the results. This claim states that memory interference creates planning difficulty, and thus it suffers from the same inability to explain why the /t/ and the vowel of to display different patterns of lengthening as seen above. While it does predict that the use of the prepositional alternant should be more likely in mid-interpretability datives where the arguments overlap both in animacy and in definiteness than in high-interpretability datives where the arguments overlap only in animacy, it does not predict the observed effect of interpretability on alternant choice when the degree of overlap between the arguments is held constant (Figure 7). Furthermore, because it posits that planning difficulty is greatest when argument overlap is greatest, it falsely predicts that disfluencies should be most common and the /t/ in to should be longest in mid-interpretability datives.

The results therefore seem to suggest that there really is a role for audience design in speech, motivated by an information-based account of communication which is sensitive to interpretability. However, this conclusion cannot be definitive due to the limitations of data sparsity. Prepositional datives are rare, and datives with arguments differing in animacy (a requirement for uncertain thematic role assignment) are ever rarer. As a result, it is extremely difficult to formulate a nuanced model of interpretability and test its relationship with the nature of the realization of the critical word to without risking non-representativeness. Here, only two low-interpretability prepositional datives were identified in a database of over 1600 dative constructions, and those two datapoints were seen to exhibit a lot of influence on statistical analyses. Many partitions including those points gave rise to effects of interpretability, making it impossible to delimit the environments in which audience design is expected to occur. We cannot even be sure that the effect of interpretability on to-realization is not spurious. To attain such surety, it would be necessary to investigate more data.

Future work could conduct such an investigation on data from another corpus or from an experiment. For example, it could investigate the Switchboard telephone corpus data previously investigated by Tily et al. (2009) and Kuperman and Bresnan (2012), which may be likely to show a greater effect of interpretability to counter the potential for noise and lack of visual signals associated with telephone conversations. Alternatively, further work could investigate data obtained from a communicative task experiment where the speaker retells a story to a listener or directs a listener to perform a task. Both approaches have their downsides: data from another corpus is likely to face the same sparsity issues as data from the present corpus, and data from an experiment may not be representative of spontaneous speech. The best approach would therefore be to combine both forms of investigation.

---

When the arguments are identical in formal properties, interpretability is typically mid-level due to at-chance thematic role assignment probability, while planning difficulty due to memory interference is maximized. Conversely, when the arguments differ in formal properties, interference-based planning difficulty is low, but interpretability may be high or low depending on whether the thematic roles are assigned in an expected or an unexpected way.
8 Conclusion

In this paper, I have explored the hypothesis that speakers avoid interpretation ambiguities by strengthening phonetic cues to critical words that exclusively support the interpretation they intend to convey when it could not confidently be inferred by a listener who didn’t hear those critical words. This exploration has been conducted through the English prepositional dative alternant, where failure to hear the critical word to could lead a listener to falsely interpret the utterance as a double-object dative alternant with the reverse assignment of thematic roles. I have grounded the hypothesis in theories of information-based communication and developed a Bayesian model for interpretability, which measures the speaker’s estimate of the probability of a listener inferring the correct interpretation of a prepositional dative without the to. A corpus study showed that speakers are more likely to use the prepositional alternant as interpretability decreases and that, in certain environments (that are explicable from an information-based standpoint), the /t/ in to in prepositional datives lengthens as interpretability decreases. These results are inconsistent with approaches that attribute a speaker’s performance solely to her own planning without considering the informativity of components of the linguistic signal to the listener. Instead, they suggest that speakers exhibit audience design, modifying the linguistic signal based on considerations of ease of interpretability for the listener. While not definitive due to limitations of data sparsity, these results nevertheless promote interpretability and audience design in an information-based framework as interesting areas of further research and beg for such research to be conducted.

References


Appendix A: Derivation of the interpretation probability formula

Suppose you observe some evidence, $E$, and want to assess how likely it is that a certain hypothesis, $H$, gave rise to it. This likelihood can be measured by the belief in the hypothesis given the evidence, $P(H|E)$. Bayes’ theorem (11) allows that probability to be estimated by a combination of your prior belief in the hypothesis before encountering the evidence, $P(H)$, the probability that the evidence would be generated under the hypothesis, $P(E|H)$, and the probability that the evidence would be generated under any hypothesis, $P(E)$. For example, how strongly I believe that it has rained given that the ground is wet is a function of my beliefs of how likely it is to rain usually, how likely it is that the ground should be wet after rain, and how likely it is that the ground should be wet usually (whether it has been raining or not).

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} \quad (11)$$
Bayes’ theorem also holds in cases where a condition \( C \) holds of the world of reference (i.e. holds of both the evidence and the hypothesis); for example, my belief in the likelihood of it having rained given the wetness of the ground may be conditioned on where I live, as that influences both the likelihood of it raining and the likelihood of the ground being wet. The degree of belief in the hypothesis given the evidence and the condition is denoted by \( P(H|E,C) \). If the condition \( C \) is fixed within the world of reference, we may use the notation \( P_C(X) \) to denote the probability of the event \( X \) in this world and rewrite (11) as (12):

\[
P_C(H|E) = \frac{P_C(E|H) \times P_C(H)}{P_C(E)}
\]

By the Law of Total Probability, the denominator in (12) may be expanded as in (13), capturing the notion that the probability of the evidence independent of any hypothesis is equivalent to the sum of the support for the evidence offered by all possible hypotheses \( H' \):

\[
P_C(H|E) = \frac{P_C(E|H) \times P_C(H)}{\sum_{H'} [P_C(E|H') \times P_C(H')]} \tag{13}
\]

Using this notation and the notion that the verb and arguments uttered by a speaker are fixed under all possible orderings of the arguments and assignments of thematic roles, the interpretation probability formula follows simply:

\[
P(\Theta = \theta|v,w,o) = P^{v,w}(\Theta = \theta|O = o)
\]

\[
= \frac{P^{v,w}(O = o|\Theta = \theta) \times P^{v,w}(\Theta = \theta)}{\sum_{\theta'} [P^{v,w}(O = o|\Theta = \theta') \times P^{v,w}(\Theta = \theta')]} \quad \text{by (13)}
\]

\[
= \frac{P(O = o|v,w,\theta) \times P(\Theta = \theta|v,w)}{\sum_{\theta'} [P(O = o|v,w,\theta') \times P(\Theta = \theta'|v,w)]} \quad \text{rewriting}
\]

Appendix B: Comparison of the syntactic alternant choice model to previous models

The fixed effects in the syntax alternant choice model shown in Table 7 agree with those seen in previous models of the dative alternation (Bresnan et al., 2007; Bresnan and Hay, 2008; Bresnan and Ford, 2010): there is a general alignment pattern whereby short, pronominal, definite and animate arguments are ordered before long, non-pronominal, indefinite and inanimate ones. However, there are also differences to previous results; namely, theme number and structural persistence were not found to have significant effects on alternant choice here, despite having such effects in Bresnan et al.’s (2007) original model for American English\(^{30}\), and speaker year of birth was

\(^{30}\)However, it should be noted that the effects of theme number and structural persistence were not strictly statistically significant in Bresnan and Ford’s (2010) revised model for American English (based on their Table 2), which is more similar to the present model in terms of variable operationalization and coding than Bresnan et al.’s (2007) original model is.
also found not to be a significant predictor\footnote{There was a near-significant ($p < 0.1$) quadratic effect of speaker year of birth here, which appeared to be driven by overfitting to peculiarities in the Intermediate Archives data as it reduced cross-validation accuracy while increasing full model accuracy. Because of this overfitting, and because the effect was not strictly significant anyway, speaker year of birth was excluded from the final model.} of alternant choice, despite having been shown to have an effect on alternant choice with the verb give in an earlier study of New Zealand English (Bresnan and Hay, 2008). The similarities are still greater than the differences, though; as Figure 9 shows, the significant effect sizes in Bresnan and Ford’s (2010) model for American English and those in the present model agree in terms of both relative and absolute magnitudes (with a possible exception for theme pronominality, which appears to have a greater effect in American English than in New Zealand English).

Figure 10 illustrates the relative importance of each predictor in the model in terms of how much explanatory value that predictor adds above and beyond that of the other predictors, measured by the decrease in model goodness-of-fit (increase in $-2 \times \log$-likelihood) when the predictor is removed from the full model. The observed importance hierarchy of predictors agrees with that seen in American English dative constructions modeled by Bresnan and Ford (2010). While this hierarchy disagrees with that seen in Bresnan and Hay’s (2008) study of datives in New Zealand English, where the pronominality of the theme had the most explanatory value, this discrepancy can be explained by the fact that their study only considered the verb give, for which many of the constructions contain pronominal themes in the form it, and thus may not have captured the theme pronominality effect in a representative way.

The verb sense biases captured by the random effect in the syntactic alternant choice model, shown in Table 8, also agree with those seen in previous research. The relative order within the verb senses in terms of bias strength and direction is consistent with that reported by Bresnan and Ford (2010) for American English data (Spearman’s $\rho = 0.59$, $p < 0.0001$), indicating that the verb sense biases of the model presented here are qualitatively similar to those reported previously. Figure 11 plots the verb sense biases in the two models against each other, illustrating their similarity but also showing that the biases associated with a few verb senses in New Zealand English have opposing directions to the corresponding biases in American English (25% of points fall in the second and fourth quadrants). Crucially, however, the verb senses with different biases in the different varieties tend to have small biases in one of the two varieties, indicating that the varietal differences do not create strong opposition. Additionally, these differences may simply be due to errors in the estimation of biases, since a number of the verb senses occur in very few tokens (at least in the data presented here).
Figure 9: Comparison of estimated model coefficients in Bresnan and Ford’s (2010) model for American English and the present model for New Zealand English. Only significant effects are shown (the same effects are significant in both models). Whiskers show 95% confidence intervals for the estimate in each model.

Figure 10: Decrease in model goodness-of-fit when each predictor is removed from the full model.
Appendix C: Back-transformation of regression data for visualization

In this section, I describe the process by which I back-transformed the regression model output into natural millisecond units and converted the original datapoints to the same scale. For concreteness, I will base the example on a model predicting /t/-duration, with the phonological environment partitioned into segments which are perceptually confusable with /t/ and those that are not. The procedure for other models is analogous.

For the jth datapoint in the data used to fit the model, let $d_j$ denote the square-root rate-normalized /t/-duration, $i_j$ the log of the interpretability, $p_j$ the log of the predictability of to from the preceding word, and $f_j$ the log of the predictability of to from the following word, and let $e_j$ be a binary indicator variable which takes the value 0 if the preceding segment is confusable with /t/ and the value 1 if it is not. Additionally, denote the mean of each continuous variable $x$ across all datapoints by $\bar{x}$. The fit model expresses a duration as determined in part by each of the predictors in the model:

$$d_j = \beta_0 + (\beta_i + \beta_i e_j)(i_j - \bar{i}) + \beta_p (p_j - \bar{p}) + \beta_f (f_j - \bar{f}) + \beta_e e_j + \epsilon_j$$

(14)

where $\beta_0$ is the intercept or baseline duration, $\beta_x$ is the coefficient estimate for predictor $x$, and $\epsilon_j$ is the jth residual, representing the error that the model makes when trying to predict $d_j$. 

Figure 11: Comparison of verb sense biases in NZ and US English.
From this expression, the duration for an unseen datapoint can be predicted via (15).

\[ \hat{d} = \beta_0 + (\beta_i + \beta_e e) (i - \bar{i}) + \beta_p (p - \bar{p}) + \beta_f (f - \bar{f}) + \beta_e e \]  

(15)

The regression line summarizing such the partial effect of interpretability (and its interaction with phonological environment) on such predictions is given by the same equation with the predictability variables assumed to take on their average values, so that the \( \beta_p \) and \( \beta_f \) terms disappear:

\[ \hat{d}_{\text{part}} = \beta_0 + (\beta_i + \beta_e e) (i - \bar{i}) + \beta_e e \]  

(16)

Since \( \hat{d}_{\text{part}} \) estimates the square root of the rate-normalized /t/-duration, we can estimate the actual rate-normalized /t/-duration \( \hat{D}_r \) by squaring, as in (17).

\[ \hat{D}_r = [\beta_0 + (\beta_i + \beta_e e) (i - \bar{i}) + \beta_e e]^2 \]  

(17)

Now \( \hat{D}_r \) represents the estimated equivalent number of syllables that the /t/ lasts for at the current rate of speech. From this, we can derive an estimate of the /t/-duration in milliseconds by assuming that the speaker is talking at the average speech rate \( \bar{r} \) and dividing this value out, as in (18).

\[ \hat{D} = \frac{[\beta_0 + (\beta_i + \beta_e e) (i - \bar{i}) + \beta_e e]^2}{\bar{r}} \]  

(18)

The log-interpretability values may also be back-transformed to the original interpretability values \( I \) by substituting \( i = \log I \) in (18); the regression lines in Section 5 plot \( \hat{D} \) against \( I \). Confidence interval lines are back-transformed in the same way.

The original datapoints can be plotted on the same scale by including their residuals in the numerator of (18), as shown in (19). This normalizes for predictability and then calculates the expected duration (in milliseconds) for a speaker speaking at the average rate, given the corresponding interpretability and phonological environment.

\[ \hat{D}_{ij} = \frac{[\beta_0 + (\beta_i + \beta_e e_j) (i_j - \bar{i}) + \beta_e e_j + e_j]^2}{\bar{r}} \]  

(19)