Formal grammar, usage probabilities, and
English tensed auxiliary contraction*

DRAFT

Joan Bresnan

*Acknowledgements: This paper would not exist were it not for the generous support of Adams Bodomo for my invitation to speak at the LFG18 Conference on July 17–19, 2018 at the University of Vienna. I am also grateful to Arto Anttila, Joan Bybee, Janet Pierrehumbert, and Geoff Pullum for providing critical comments on drafts of the write up which led to substantial improvements in the present paper, though of course they are blameless for any errors remaining.
At first sight, formal theories of grammar and usage-based linguistics appear completely opposed in their fundamental assumptions (Diessel 2007). If grammatical structures emerge from language use and are constantly changing through psychological processing, as usage-based grammar claims, how can they be formalized as a closed and stable system, independent of use? As one step toward answering this question, the present study will show that Wescoat’s (2002, 2005) theory of lexical sharing in LFG provides a way to combine major findings from formal and usage-based lines of research on tensed auxiliary contraction in English. The results potentially extend to lexicalist theories of syntax more generally.

Tensed auxiliary contraction in English is particularly interesting because it appears to create linguistic units in the syntax that are not also syntactic or semantic units. For example, the contraction law’s in the sentence my other brother in law’s Arab (authentic example from the Buckeye corpus, Pitt et al. 2005) is not a compositional component in the semantics of the sentence, even though the clitic auxiliary ’s provides a coda of the open syllable of law that selects the voiced variant [z] (in contrast to his wife’s a teacher, which selects [s]). Nor does the contraction behave as a syntactic constituent: *Who’s do you think coming? vs. Who do you think is coming? (Anderson 2008:174). This interrogative construction otherwise allows larger syntactic constituents to appear with the focused phrase, as in At what time do you think she’s coming?. For these reasons tensed auxiliary contraction has long been treated in formal linguistic frameworks as simple cliticization (Zwicky 1977), a phonological grouping of two adjacent non-constituent words in the syntax, belonging to the surface syntactic phrasings of metrical and prosodic phonology (Selkirk 1984, 1996; Inkelas & Zec 1993; Anderson 2008; Anttila 2017)—purely post-lexical phonological processes.

Yet that is far from the whole story: a number of researchers have pointed out morphophonological properties of the most common auxiliary contractions that are signs of the contracted forms being lexically stored (Kaisse 1985; A. Spencer 1991; Bybee & Scheibman 1999; Scheibman 2000; Wescoat 2005, Bybee 2006). And usage statistics show that the probability that words will be

---

1The term “auxiliary” includes the copula in the present study, because it shares the syntactic properties that distinguish auxiliary verbs from main verbs. These include not placement, n’t contraction, and subject-verb inversion: She is not sleeping/sleepy vs. *She sleeps not; She isn’t sleeping/sleepy vs. *She sleepsn’t; and Is she sleeping/sleepy? vs. *Sleeps she? See Huddleston & Pullum (2002) for discussion of the full set of distinguishing properties of auxiliary verbs.
adjacent in naturally occurring speech determines their “degree of fusion” into lexical units (Bybee & Scheibman 1999, Scheibman 2000, Bybee 2002) and their likelihood of contraction (Krug 1998; Frank & Jaeger 2008; J. Spencer 2014; Barth & Kapatsinski 2017).

What appears to be needed to explain fully the properties of tensed auxiliary contractions is a theory of their representations that simultaneously accounts for their syntactic non-constituency and adjacency constraints, their lexical morphophonology, their prosodic and metrical phonology as well as the effects of usage probability on their degree of morphophonological fusion and their likelihood of contraction. In other words what is needed is a theory that can account for the combined findings of formal and usage-based studies of tensed verb contraction.

Unfortunately, although tensed auxiliary contraction in English is one of the empirical domains that have attracted research in both formal and usage-based theories of grammar, the two lines of research have proceeded mostly independently and have thus failed to provide a full answer to the deeper questions contraction poses. “Formal” research on English auxiliary contraction includes analyses in various systems of generative grammar (such as Zwicky 1970; Baker 1971; Bresnan 1971; Kaisse 1983, 1985; Zwicky & Pullum 1983; Selkirk 1984, 1996; Klavans 1985; Inkelas & Zec 1993; Wilder 1997; Sadler 1998; Barron 1998; Bender & Sag 2001; Wescoat 2002, 2005; Anderson 2008; Anttila 2017). “Usage-based” research on English auxiliary contraction has included earlier work examining frequency effects on contractions (Krug 1998; Bybee & Scheibman 1999; Scheibman 2000, Bybee 2001, Bybee 2002) and more recent corpus studies of the probabilities of actual uses of contraction, employing quantitative methods such as statistical modeling of data using communication-theoretic measures such as information content (for example, Barth 2011; Frank & Jaeger 2008; Spencer 2014, Barth & Kapatsinski 2017). Sociolinguistic research on the topic in the Labovian tradition has generally adopted both quantitative methods and the representational basis of generative grammar, usually with the primary focus on relating the grammar of the copula to social factors (Labov 1969; McElhinny 1993; Rickford et al. 1991; MacKenzie 2012, 2013).

The present study of tensed auxiliary contraction proposes that the formal syntactic theory of lexical sharing in LFG (Wescoat 2002, 2005), embedded in a hybrid exemplar-dynamic model of the mental lexicon (Pierrehumbert 2001, 2002, 2006) can provide the necessary combined approach. Lexical sharing in LFG was originally designed to account for narrowly defined types of cases where lexical units do not match constituent structure units, such as contractions of
preposition-determiner combinations (witness German *zum, am, im, ins*, etc. and French *du, au, des, aux*, etc.), and contractions of simple clitics like English tensed auxiliary contractions, the subject of the present study. However, as the present study shows, it naturally extends to the lexicalization of multi-word sequences in larger constructions. While the original work solved formal representational problems in theoretical syntax, it—like most work in formal grammar—ignored the role of usage probabilities in lexicalization. On the other hand, usage-based linguistic studies of the same phenomena have seldom presented fully articulated proposals for their syntactic representations. The present study therefore contributes to both formal and usage-based lines of research.

The first three sections below outline some of the main findings of usage-based linguistics on tensed auxiliary contractions and show how they are explained theoretically. The next three sections present the theory of lexical sharing and show how it captures the main findings of formal research on tensed auxiliary contraction. The two sections following those show how these two approaches naturally fit together and yield further empirical consequences. The final sections summarize the main findings of this study and then sequences then discuss some of its broader implications for formal grammar.

### 1 Usage and phonetic reduction

A major finding of usage-based linguistics is that more probable words and multi-word expressions are phonetically more reduced and become lexically stored (Bybee 2001, 2006; Bybee & Hopper 2001; Pierrehumbert 2001, 2002, 2006; Seyfarth 2014; Sóskuthy & Hay 2017). For example, Bybee & Scheibman (1999) show that in *don’t* contraction, the reduction process is most advanced with the most frequent context words and the reduced multiword forms have accrued additional pragmatic functions along with the changes in form, suggesting their lexical storage. These are typical effects of lexicalization: when composite items are lexically stored as wholes, they begin to acquire their own usage profiles and drift in their grammatical and semantic properties from their constituent elements.

Bybee & Scheibman (1999) collected and transcribed tokens of *don’t* from about three hours and 45 minutes of “naturally occurring conversations.” In Table 1, which gives excerpts from Bybee & Scheibman (1999:581–582), the words of the left and right contexts of *don’t* are ordered by frequency from top
to bottom. Thus pronouns, as preceding contexts of don’t, are far more frequent than lexical NPs and among the pronouns, I is the most frequent. As following contexts of don’t, the verbs know and think are the most frequent. The extent of phonetic reduction increases from left to right: the final stop deletes, the initial stop becomes a flap and then also deletes, and the vowel reduces, so that ultimately don’t is pronounced as a nasalized schwa. As the table shows, don’t is more highly reduced phonetically in the most frequent contexts I and _ know, _ think, than in all others.

<table>
<thead>
<tr>
<th>Preceding</th>
<th>[dɒt, dɵ]</th>
<th>[rʊt, rɵ]</th>
<th>[r̩]</th>
<th>ʃ</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>16</td>
<td>22</td>
<td>38</td>
<td>12</td>
<td>88</td>
</tr>
<tr>
<td>you</td>
<td>7</td>
<td>7</td>
<td></td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>we</td>
<td>2</td>
<td>6</td>
<td></td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>they</td>
<td>1</td>
<td>3</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>lexical NP</td>
<td>5</td>
<td></td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Following</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>know</td>
<td>2</td>
<td>8</td>
<td>24</td>
<td>5</td>
<td>39</td>
</tr>
<tr>
<td>think</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>have</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>have to</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>want</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>see</td>
<td>3</td>
<td>1</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Don’t variants by type of preceding and following item in data from Bybee & Scheibman 1999:581–582

According to Bybee & Scheibman (1999), these developments arise when frequent motor repetition in articulation becomes automatized, the automatization of pronunciation leads to blurring of word and morpheme boundaries and compression of entire multiword units, and over time the result becomes a new lexically stored unit, which separately accrues its own characteristics of form and function. Lexicalization occurs because “lexical storage is highly affected by language use, such that high-frequency forms have stronger lexical representation than low-frequency forms” (Bybee & Scheibman 1999:583). As shown in Table 2 the reduced-vowel variants of don’t in I don’t know contrast overwhelmingly with the full-vowel variants in expressing special pragmatic functions of “indicating speaker uncertainty and mitigating polite disagreement in conversation” (Bybee & Scheibman 1999:587) in addition to the literal lexical sense.2

2Applying a one-sided Fisher exact test to Table 2 to ascertain whether the odds ratio
Table 2: Full-vowel and reduced-vowel variants of don’t by lexical versus pragmatic function in data from Bybee & Scheibman (1999:587)

<table>
<thead>
<tr>
<th></th>
<th>Full vowel</th>
<th>Schwa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical sense</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>Pragmatic function</td>
<td>1</td>
<td>17</td>
</tr>
</tbody>
</table>

2 Usage and syntactic contraction

Another major finding is that the syntactic contraction, or cliticization, of word sequences is most advanced among the sequences with the highest usage probabilities. Consider tensed auxiliary contraction, which occurs when a specific set of tense-bearing auxiliary verbs, including is, are, am, has, have, will, and would, lose all but their final segments, orthographically represented as ’s, ’re, ’m, ’s, ’ve, ’ll, and ’d, and form a unit with the immediately preceding word, called the HOST.

Although the influential early formal analysis of Labov (1969) treats the contracted verb forms as phonological reductions of the full uncontracted forms, many subsequent phonological analyses hold that synchronically, the contracted forms are are allomorphs of the full forms (Kaisse 1985; Inkelas 1991; Inkelas & Zec 1993; Anderson 2008; Mackenzie 2012, 2013). Evidence for analyzing contracting auxiliaries as morphological variants rather than phonological reductions or rapid-speech effects includes (1) the fact that there are grammatical differences between the contracted and full forms: e.g. *there’s three men outside* vs. *there is three men outside* (see Dixon 1977, Nathan 1981, Sparks 1984, Kaisse 1985); (2) that phonological rules that delete the onsets and schwas of specific auxiliaries cannot be assimilated to post-lexical “rapid-speech phenomena such as deletion of flaps, coalescence of vowels etc.” (Kaisse 1983:95); (3) that the phonology of specific contractions cannot be assimilated to function-word reduction in general (Kaisse 1985); and (4) that speech rate is not predictive of auxiliary contraction in spoken corpus data (Frank & Jaeger 2008). It is also worth noting that auxiliary contraction cannot simply be assimilated to casual speech (McElhinny 1993:376): in style-shifting among white speakers, *is* contraction occurred 79% of the time in casual speech (in group interviews) and 87% of the time in careful speech (in single interviews) (Labov 1969:730–731).

of vowel reduction co-occurring with the pragmatic function is reliably greater than 1, as predicted, yields $p$-value $= 0.02545$. 

5
A usage-based corpus study of tensed auxiliary contraction in “spoken mainstream British English” by Krug (1998) finds that the contraction of tensed auxiliary verbs (e.g. I’ve, he’s, we’ll) varies directly with the bigram probability (“string frequency”) of the subject and the auxiliary. Even where the preceding phonological contexts are similar—open monosyllables ending in tensed vowels in I’ve, you’ve, we’ve, they’ve, who’ve—the bigram probability directly correlates with the proportions of contractions.\(^3\)

Recent work on several other varieties of spoken English has confirmed the basic finding that probabilistic measures derived from frequencies of use of hosts and auxiliaries correlate with the likelihood of contraction (Frank & Jaeger 2008, Barth 2011, Barth & Kapatsinski 2017, J. Spencer 2014). These works employ counts of the frequency of use of host-auxiliary sequences to estimate their probabilities, from which they calculate transition probabilities, conditional probabilities, information content, and related measures.

For example, J. Spencer (2014) collected variable present tense aux be tokens for the most frequent pronoun hosts in the Buckeye corpus (Pitt et al. 2005), and the author of the present study did the same for the present tense aux have tokens. Each scaled the host+verb bigram probabilities by the probabilities of the inflected verbs (differing by person and number: am/m, is/s, and are/re; have/ve or has/s), to derive

$$\log\left(\frac{1}{P(\text{host}|\text{verb})}\right),$$

expressing the information content of the host in the context of the verb. Against this quantity the log odds of contraction,

$$\log\left(\frac{P(\text{contracted})}{P(\text{uncontracted})}\right),$$

is plotted in Figure 1.

\(^3\)In a corpus study of contractions in the Switchboard corpus (Godfrey & Holliman 1997), MacKenzie (2012:130,149–155) finds that the frequency effect on contraction “does hold for the extreme ends of the frequency scale (i.e., the most and least frequent host/auxiliary combinations do contract at a high and a low rate, respectively), but that the string frequency/contraction connection does not hold to any degree of granularity in the middle,” concluding that “the attested pronoun-specific effects on short allomorph selection cannot be explained by string frequency alone.” Her results are based on (estimated) raw string frequencies, as are the findings of Krug (1998). The findings discussed in Section 8 of the present study and those of Frank & Jaeger (2008), Barth (2011), Barth & Kapatsinski (2017), and J. Spencer (2014)) support effects of usage probability on contraction throughout the frequency range.
Figure 1: Relation between the information content of the most frequent hosts before the verb form and the log odds of contraction in the Buckeye corpus. The *have* and *be* datasets were separately collected and are respectively plotted with magenta and cyan dots.

Figure 1 clearly shows a strong inverse relation between information content of the most frequent pronoun hosts and the log likelihood of contraction: the first person singular pronoun *I* has the least information content before the first person singular verb form *am*, and that sequence has the highest log likelihood of contraction. As information content increases from left to right, the log likelihood of contraction shows almost linear decrease for present-tense forms of both *be* and *have.*
3 The mental lexicon

What explains the close relation between usage probability and contraction? Krug (1998:305) hypothesizes that the word, or sequence of words, in subject-auxiliary contractions is stored in the mental lexicon, which responds dynamically to usage probabilities as proposed by Bybee (1985:117):

"Each time a word is heard and produced it leaves a slight trace on the [mental] lexicon, it increases its lexical strength."

Pierrehumbert’s (2001, 2002, 2006) exemplar-dynamics model fleshes out this concept of the mental lexicon: it consists essentially of a map of the perceptual space and a set of labels, or structural descriptions, over this map. Long-term memory traces are located in the perceptual space. Each exemplar has an associated strength or resting activation; exemplars of frequent recent experiences have higher resting activation levels than those of infrequent and temporally remote experiences.

Figure 2 provides a simplified visualization of tensed auxiliary contractions in this model. The labels you, you’re, and are with their varying pronunciations stand for (partial) ‘lexical entries’ in traditional linguistic terminology and correspond to structural descriptions at several levels, not shown (see Wright et al. 2005, German et al. 2006). Each entry maps onto a matching set of remembered instances of its utterance—the memory traces (or exemplars). The visualization is simplified to show only varying pronunciations of remembered instances; it omits links to further grammatical, pragmatic, semantic, and social information. Fresh experiences and memory decay lead to continual updating of the entries in the mental lexicon.

The mental lexicon implicitly encodes bigram probabilities and information content as activation levels of the various words and multi-word structures that are stored there. The crucial connection between high-probability/low-informativity host-auxiliary bigrams and higher incidences of contraction in speech production is then straightforward:

- High-probability/low-informativity word sequences are produced more often than low-probability/high-informativity sequences, so their contractions leave denser clouds of memory traces.

---

4 The theoretical types of frequency effects generated by the model depend on the parameter range for memory decay and are broader than discussed here.
Figure 2: Visualization of tensed auxiliary contractions in an exemplar-dynamic model of the mental lexicon (Pierrehumbert 2001, 2002, 2006), which includes memory traces of speech events of varying activation levels (not represented).

- In language production an exemplar of a label is randomly sampled from its cloud as the goal of production; in this way the denser, more highly activated clouds bias the speaker toward the higher-probability/low-informativity outputs.\(^5\)

If highly probable and relatively uninformative contractions are lexically stored with phonetic detail, they should accrue their phonetic reductions and coarticulations from the repetitive automatization of pronunciation as part of their long-term representations. As Bybee (2006:723) puts it, “Frequent phrases such as I don’t know, I don’t think, and many others show phonological reduction in excess of that which could be attributed to on-line processes, such as that evident in other tokens of don’t, as in I don’t inhale, indicating that such reduction has accumulated in representation.” There is evidence that fits this expectation.

Wescoat (2005:471) gives various examples of “morphophonological idiosyncrasies” among tensed auxiliary contractions, shown in Table 3. One of them is that “I [ai] may be pronounced [a], but only in association with ’ll (will),

\(^5\)The relation between production and perception assumed here is obviously simplified. Further, there is evidence that word frequency effects vary with the production or perception task (Harmon & Kapatsinski 2017) and ‘word prevalence’—how many different people use a word—may also contribute to frequency effects on lexical decision times (Brysbaert et al. 2016).
yielding [əl]; moreover you may become [jʊ], but only when followed by ’re (are), resulting in you’re [jʊə].” Thus the reduced pronunciations are specific to individual pronoun-auxiliary sequences. He emphasizes that these pronunciations are not fast-speech phenomena: *I’ll [əl] and you’re [jʊə] “may be heavily stressed and elongated.” In other words, their pronunciations are not merely on-line contextual adjustments to the phonology of rapid connected speech.

<table>
<thead>
<tr>
<th>Pronunciation</th>
<th>Reduced Pronunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>I’ll</em></td>
<td>[əl/əl]</td>
</tr>
<tr>
<td><em>I’m</em></td>
<td>[aɪm/*aɪm]</td>
</tr>
<tr>
<td><em>I’ve</em></td>
<td>[aɪv/*aɪv]</td>
</tr>
<tr>
<td><em>you’ll</em></td>
<td>[jʊːl/*jʊəl]</td>
</tr>
<tr>
<td><em>you’re</em></td>
<td>[jʊə/*jʊə]</td>
</tr>
<tr>
<td><em>you’ve</em></td>
<td>[jʊv/*jʊv]</td>
</tr>
</tbody>
</table>

Table 3: Contrasting contraction-specific pronunciations from Wescoat (2005)

Diachronically, these pronunciations could theoretically derive from such on-line contextual adjustments (for example, the velarization of /l/ in *will* and the lowering and backing of immediately preceding unstressed vowels, yielding *we’ll* [wɪl] in contrast to the rhyming but stressed proper noun *the Cree’ll* [*kʊɪl*]). But the retention of the reduced pronunciations of specific words even in slow or emphatic speech shows that synchronically, their distribution does not match that of on-line contextual adjustments to the phonology of rapid connected speech. It rather supports lexical representation of the reduced variants. The simplest account is that synchronically they are lexically stored allomorphs of the contracting pronouns.

Along the same lines, Piantadosi et al. (2011) show from a cross-language corpus study that information content is an important predictor of orthographic word length (more so than frequency itself), across lexicons from a variety of languages:

One likely mechanism for how the lexicon comes to reflect predictability is that information content is known to influence the amount of time speakers take to pronounce a word: words and phones are given shorter pronunciations in contexts in which they are highly predictable or convey less information [references omitted]. If these production patterns are lexicalized, word length will come to depend on average informativeness.

The Bybee-Pierrehumbert theory of the mental lexicon provides an explicit model of the lexicalization of production patterns in which more probable (less

---

Here he describes his own speech, but notes that Sweet (1890:25) also reports this pronunciation of you’re, and it is shared by the present author as well.
informative) words become reduced (shorter). Seyfarth (2014) discusses this and possible alternative models of the effects of informativity, or average contextual predictability, on lexicalization of words’ durations. All of the alternatives he discusses but one assume with Bybee and Pierrehumbert that both reduced forms and their probabilities of use are lexically stored; hence, all of these alternatives are broadly consistent with the hybrid formal/usage-based approach described here, and may be regarded as different implementations of the fundamental usage-based insight connecting lexicalization with probability and reduction. The exemplar-dynamic implementation is adopted here for its intuitive expository value, but alternative implementations of the usage-based insights could be adopted as well.

4 Lexical sharing and restricted contractions

The theory of lexical sharing in LFG (Wescoat 2002, 2005) provides a formal analysis of tensed auxiliary contractions in English that turns out to be highly compatible with usage-based findings for these phenomena, and is also broadly extendable. In this theory, morphological and phonological units do not have to be associated with just one terminal category node in the syntactic structure, but can be shared between two linearly adjacent terminal category nodes. Figure 3 provides an illustration of the idea.

In Figure 3 the arrows pointing to words represent a formal mapping from syntactic constituent structures (c-structures in LFG) to the lexical items that instantiate them. As usual in LFG the c-structure represents the “surface” syntactic groupings of words, while the “deeper” relations and dependencies are provided in a parallel functional structure (f-structure) that bears many similarities to dependency grammar graphs (Mel’čuk 1988, Bresnan 2016). The surface words themselves provide most of the global functional information in the form of relational features that give rise to descriptions of the f-structure context of the word. Language-particular c-structure configurations provide what structural information about linguistic functions there may be in a given

\[\text{\cite{Wescoat2002, Wescoat2005} for discussion of a distributed connectionist alternative model.}\]

\[\text{\cite{BybeeMcClelland2005} for discussion of a distributed connectionist alternative model.}\]

\[\text{\cite{BybeeScheibman1999} for discussion of a distributed connectionist alternative model.}\]
language, which in the case of configurational languages like English is fairly redundant (Bresnan et al. 2015).

Wescoat (2005) initially applies lexical sharing to a subset of “restricted” contractions which have “morphophonological idiosyncrasies” and “functional restrictions”: “The nonsyllabic contractions of am, are, have, and will (and for some speakers, had and would) are attached to pronouns and wh-words IN THE LEXICON” (Wescoat 2005:482). In the lexicon these restricted contractions are associated with adjacent syntactic terminal categories and may specify item-specific phonology and functional restrictions, as illustrated in (1). In (1) the lexical entry for you’re specifies the pronunciations indicated and shows that the contraction is lexically shared by the sequence of adjacent categories D and I.

The ‘down’ arrows in (1) are standard LFG metavariables which give rise to functional structures when instantiated in the syntactic context of a particular sentence, phrase, or fragment of language. The double down arrow ↓ is a special metavariable defined by Wescoat (2005) to refer to the f-structure of the lexical exponent of a category. In the case of a contraction like you’re in (1), which is the lexical exponent of two adjacent categories, the double down arrow allows properties of the f-structure of the contraction as a whole to be specified in addition to the standard properties of the f-structures of its atomic D and I elements. Specifically, the equation ↓ = ↓ identifies the functional structure of the host you with that of the entire contraction, while the equation (↓ subj) =c ↓ imposes the constraint that the host must be the subject of the auxiliary ‘re. To be more precise, it specifies that the f-structure of the contraction (which is identified with that of the atomic host D) must be the value of the subj function of the atomic auxiliary I f-structure.

The particular category names are not important; here Wescoat follows the c-structure.
(1) Lexical entries for the structure in Figure 3:

you’re [juːr/ʃuːr/ʃʊə] ← D
  \(↓\) PRED = ‘PRO’
  \(↓\) PERS = 2
  \(down\) = \(down\)

I
  \(↓\) TENSE = PRES
  \(↓\) SUBJ NUM = PL
  \(↓\) SUBJ = c \(down\)

going [gʊʊŋ] ← V
  \(↓\) PRED = ‘GO\{\(\down\)SUBJ\}\’
  \(↓\) ASP = PROG
  \(down\) = \(down\)

Given these lexical entries and the c-structure shown in Figure 3, Wescoat shows that the correct f-structures follow from general principles of structure-function mapping (Bresnan 2001, 103; Bresnan et al. 2015). These are visualized in Figure 4; the linking arrows show how the global f-structure corresponds to the c-structure phrases of which D and I are head and co-head, lexically sharing the contraction you’re which provides their substantive features. (See Wescoat (2005) for more details.)

This analysis captures the main morphophonological and syntactic findings of formal analyses of restricted contractions (A. Spencer 1991; Barron 1998; theory outlined by Bresnan (2001) and Bresnan et al. (2015), but any appropriate category labels will do.
Sadler 1998; Bender & Sag 2001; Wescoat 2002, 2005). For example, as Zwicky (1970:332) initially observes, restricted contractions generally require a pronoun subject. In (2)a the asyllabic pronunciations of ‘re with you are not possible, indicating that contraction does not occur when the host is not the subject (but only a dependent of the subject). Likewise, the laxed monosyllabic pronunciations of contracted we’re, we’ll with the pronoun host we in (2)b are not possible with the rhyming noun Cree in (2)c.

(2) a. [The people beside you] ‘re going. [juː.ː/*juː.l/*jʊl/*jɔl]
   
   b. We’re a big group. [wu]  
   We’ll win. [wil]
   
   c. The Cree’re a big group. [kii.ː/*kiii]  
   The Cree’ll win. [kii.ː/*kiii]

To these well known types of cases where a non-subject pronoun fails to host contraction of an adjacent restricted auxiliary may be added (3)a,b:

(3) a. [The people who want to meet you] ‘re going. [juː.ː/*juː.l/*jʊl/*jɔl]
   
   b. Which ones did they [tell you] [‘re going?] [juː.ː/*juː.l/*jʊl/*jɔl]

In both (3)a,b the pronoun you before the adjacent auxiliary is the object of the preceding verb and not the subject of the following auxiliary, as required in (1). Accordingly, contraction fails.\(^\text{11}\)

Wescoat broadens the analysis from pronoun subjects to include interrogatives bearing grammaticalized discourse functions (DF) in LFG, and also assumes that the auxiliary may be in C as the extended head of its clause (Bresnan 2001, 103; Bresnan et al. 2015):

(4) how’ve [hauv] ← ADV C
   \(\downarrow\text{PRED} = \text{‘HOW’}\)
   \(\downarrow = \downarrow\)
   \(\downarrow\text{TENSE} = \text{PRES}\)
   \(\downarrow\text{ASPECT} = \text{PERF}\)
   \(\downarrow\text{FOCUS} = c \downarrow\)

\(^{11}\)Note that the object you in both (3)a and (3)b is an unstressed pronoun permitting stress retraction to the verb (cf. Anttila’s 2017 stress-based theory of contraction), and that in (3)b you c-commands the auxiliary, in the sense that every maximal node that dominates it also dominates the auxiliary (cf. Kaisse’s 1985 c-command condition on contraction with pronoun hosts, pp. 55–56).
This extension allows restrictive contractions with interrogative pronouns in a parallel way, capturing the noncontracted pronunciation of 've with the noun Au, which Wescoat uses to refer to speakers of the Au language of Papua New Guinea (Simons & Fennig 2018).

(5) **How’ve you been? [hauv]**
    *The Au’ve been polled. [au.ɔv/*auv]*

Wescoat (2005) enumerates several syntactic consequences of his analysis within the LFG framework. He shows how restricted contractions have the syntactic consequences listed in (6)a–c.

(6) a. Restricted contractions cannot contract with a right conjunct:
    *[John and you | ’re going. [ju:/*ju:*/*jo/*/*jo/*]*]

b. They cannot be conjoined:
    *[^they’re and you’re | going.]

c. They permit I’ coordination:
    *I’ll [aAl/*aAl] be there on Sunday and [I’ am looking forward to seeing you]*

In addition to accounting for these syntactic restrictions, Wescoat’s (2005) theory of lexical sharing expresses both the syntactic non-constituency and the lexical morphophonology of restricted contractions.

5 Lexical sharing and unrestricted contractions

Wescoat (2005:482) proposes extending the theory of lexical sharing from restricted contractions of tensed auxiliaries to the unrestricted tensed 's contractions (and indeed to all simple clitics), but he leaves the analysis undeveloped beyond these comments:

“There is a lexical process that attaches 's [z/s/ɔz] (is or has) to a host, yielding a lexical-sharing structure; the host may be anything, the attachment of 's [z/s/ɔz] triggers no morphophonological idiosyncrasies, and no functional restrictions are involved. The lack of morphophonological and functional intricacies in no way undermines a lexical-sharing analysis.”
It is not difficult, however, to provide a lexical sharing analysis of 's contractions. (7) shows the schematic form of lexical entries of contractions of 's. It differs from the entry for you’re shown in (1) in that here the restriction (↓ subj) = c ↓ is absent and the host and its category are unspecified.

(7) Schematic form of lexical entries of contractions of 's:

\[
\begin{align*}
x's & \rightarrow X & I \\
\downarrow & \rightarrow \downarrow & (\downarrow \text{TENSE}) = \text{PRES} \\
\downarrow = \downarrow & \rightarrow (\downarrow \text{SUBJ NUM}) = \text{SG} \\
& & (\downarrow \text{SUBJ PERS}) = 3
\end{align*}
\]

The choice of the specific pronunciation of 's in (7) depends on the phonology of the host x. Morphophonologically, 's contractions undergo word-internal rules of voicing assimilation or epenthesis—or perhaps more accurately, phonologically conditioned allomorph selection among the variants [z/s/iz]—parallel to plural and tense inflections:

(8) a. **plurals:** peats ([s]), reds ([z]), losses ([az])

   b. **present tense:** bleats ([s]), shreds ([z]), tosses ([az])

c. **'s contractions:** Pete’s ([s]) here, Fred’s ([z]) here, Ross’s ([iz]) here

The contrast with arbitrary adjacent syntactic words shows that the voicing assimilation and epenthesis are word-internal effects specific to contractions with the auxiliary 's:

(9) Pete sang ([s]) and Fred sang ([s/*z])

Fred zigged ([z]) and Pete zagged ([z/*s])

Ross zig-zagged ([z/*oz])

An example of 's contraction under lexical sharing is given in Figure 5, and the lexical entry of the contraction blood’s is given in (10). Note that the lexical entry has the schematic structure in (7), which requires adjacency in c-structure between the host and auxiliary categories.

Figure 6 shows how the structure in Figure 5 corresponds to the global f-structure that results from the same principles of structure-function mapping as before. Under this theory D and NP are co-heads, just as I and VP are co-heads.
Because the f-structures of co-heads unify, the features of the NP dominating
the host N are unified with the features of the proximate demonstrative D this.

(10) Lexical entry for the contraction blood’s in Figure 5.

\[
\text{blood’s} \left[ \text{blʌdz} \right] \leftarrow \begin{array}{l}
\text{N} \\
(\downarrow \text{PRED}) = \text{‘BLOOD'} \\
\downarrow = \downarrow \\
(\downarrow \text{TENSE}) = \text{PRES} \\
(\downarrow \text{SUBJ NUM}) = \text{SG} \\
(\downarrow \text{SUBJ PERS}) = 3
\end{array}
\]

Syntactically, ’s contraction occurs freely in contexts that are closed to
contractions with ’re and the other restricted auxiliaries. Authentic spoken
examples (11)a–b from the Canterbury Corpus (Gordon et al. 2004) and (12)a–b
from the Buckeye Corpus (Pitt et al. 2005) illustrate this fact, showing ’s
contractions with noun hosts, dependents of the subject of the auxiliary, and
conjuncts.

(11) a. [the computer science department at Canterbury]’s really lousy

b. [anything to do with money]’s good

(12) a. [everybody in my family]’s mechanically inclined

b. [August September and October]’s just gorgeous

\footnote{The generalization to contractions of inverted ’s would allow C as an extended head as
well as I; see discussion of (4).}
A striking property of 's contraction, known at least since Baker (1971) and Bresnan (1971), is that 's contracts from a sentential complement across a wh-extracted subject to a superordinate verb. Examples (13)a–c are authentic examples from the web, selected with negation of the host verb and an affirmative complement in order to eliminate parenthetical readings:

(13) a. I'll tell you what I don't think's going on. \[\theta\eta ks\]

b. What I don't think's beautiful is a boy in my daughter's bedroom. \[\theta\eta ks\]

c. You can't oppose what you don’t know's happening. \[\text{nooz}\]

As (14) and Figure 7 show, the lexical sharing analysis of these cases is straightforward.

(14) Lexical entry for the contraction think's:

\[
\text{think's } \theta\eta ks \leftarrow V \downarrow \text{PRED} = \text{‘THINK( (SUBJ) (COMP))’} \quad \downarrow = \downarrow \quad \text{I} \downarrow \text{TENSE} = \text{PRES} \\
\quad \downarrow \text{SUBJ NUM} = \text{SG} \\
\quad \downarrow \text{SUBJ PERS} = 3
\]
It is particularly noteworthy that the lexical entry in (14) is the same in schematic form as that in (10), even though the resulting grammatical relations between the host noun and auxiliary are entirely reversed. To see the reversal, compare Figure 6, where the host heads a subject which is an argument of the main clause co-headed by the tensed auxiliary, to Figure 7, where the host heads the main clause and the tensed auxiliary co-heads a complement clause which is an argument of the host predicate. No special stipulations of functional annotations are required to derive the correct f-structures. Both structures satisfy the adjacency requirements of the schema for unrestricted contractions in (7) and follow from the general principles of structure-function mapping invoked by Wescoat (2005).

![Figure 7: C-structure to f-structure links for a structure using the lexical entry in (14)](image)

Furthermore, since *is* contractions are not c-structure constituents under lexical sharing, there is no danger of unwanted ‘movements’ in the lexical shar-
ing analysis of unrestricted constrictions (cf. Anderson 2008, 174):

(15) *Who’s do you think coming?
    cf. Who do you think is coming?

6 The phonological word

Nevertheless, there are restrictions on is contraction, at least for many speakers, as the examples in (16) from Kaisse (1979, 1983, 1985) and MacKenzie (2012) illustrate:

(16) a. *Speaking tonight’s our star reporter.
    b. *That John finally ate’s making Momma happy.
    c. *What I’m talking about’s the people over here and over here and across the street.

Inkelas & Zec (1993:243,245) propose a phonological explanation for such phenomena, following Sells (1993). They assume that English auxiliary clitics form a phonological word ω with a phonological word to their left: [ [ [ |ω CL ]ω]. Then they assume with Sells (1993) that certain focused syntactic constituents are obligatorily set off by a phonological or intonation phrase boundary which prevents auxiliary enclitization. Examples (17)a–c illustrate this proposal:\[13\]

(17) a. *{ speaking (tonight \{ ’s\}_ω our star reporter \}
    b. *{ that John finally (ate \{ ’s\}_ω making Momma happy \}
    c. *{ what I’m talking (about \{ ’s\}_ω the people over here . . . \}

If this phonological analysis is correct, these facts already follow from the lexical sharing analysis of tensed auxiliary contractions, given the widely shared assumption of prosodic phonologists that ALL LEXICAL WORDS ARE PHONOLOGICAL WORDS. If lexically shared contractions ( X + I )ω are phonological words, they cannot be interrupted by pauses or by the obligatory prosodic boundaries of certain focused or dislocated syntactic phrases.

\[13\]Inkelas & Zec explicitly discuss only (16)a, but their proposal appears applicable to all of (16)a–c as well as the other inversion phenomena discussed by Kaisse (1983).
In contrast, the tensed weak SYLLABIC auxiliaries are phonologically dependent on their RIGHTWARD phrasal context only (Inkelas 1991; Inkelas & Zec 1993; Selkirk 1984, 1996). Hence pauses and strong prosodic boundaries can separate them from the preceding word:

(18) a. { speaking tonight }{ is [iz] our star reporter }

b. { that John finally ate }{ is [iz] making Momma happy }

b. { what I’m talking about } { is [iz] the people over here and over here and across the street }

(19) { They—bicycle cranks, I mean—are [əi] expensive }

Under the lexical sharing analysis, both the asyllabic forms and the unstressed syllabic forms of the tensed auxiliaries have the same syntactic position at the left edge of their syntactic constituents. Both are stressless and both are metrically dependent on their complement constituents to the right. This relation is what Selkirk (1984:405) describes as “the central generalization” about auxiliary contraction:15 “only auxiliaries that would be realized as stressless in their surface context may appear in contracted form.” Phrase-final position is one context in which stress is required and both forms of stressless tensed auxiliaries are blocked:

(20) They are/*’re [ᵦeɪ.əɪ/*ᵦeɪ.əɪ/*ᵦeɪ]  

The generative literature includes constructed examples where phrase-medial sites of gaps and ellipsis block contraction (e.g. Bresnan 1973, Kaisse 1985, Inkelas & Zec 1993, Wilder 1997). But medial sites of blockage, reported to be obligatory on the basis of linguistic judgments, seem actually to be variable in usage. Selkirk (1984) and Inkelas & Zec (1993) provide cases where contraction can occur directly before medial syntactic gaps:

(21) a. I don’t know how much there’s left in the tank (Inkelas & Zec 1993:247)  
(cf. There is how much left in the tank?)

14Following Inkelas & Zec (1993), the [i] transcription in (18)a–c represents an unstressed is, even though pronunciations of the unstressed vowel may vary.

15This is also the core generalization of Labov’s (1969) analysis. Mackenzie (2012, 2013) observes the same in her corpus data.
b. *Looks as good as it’s ___ fun to play* (Selkirk 1984:443, n.25)

Attested examples in support of their contention can be found on the web:

(22) a. *Hi, Soon going to London, and I’ve got an Oystercard from last time. Is there any possibility to see how much there’s left on it and/or top up online?*

b. *So many have chimed in on Lin at this point that we’re not even sure how much there’s left to say.*

(23) a. “*But I know he’s a better runner than he’s a biker,*” Lopez said.

b. *If it’s longer than it’s wide, then it’s phallic. If it’s not longer than it’s wide, then you put it on its side. Now it’s longer than it’s wide, and it’s phallic!*

Accounting for the variability of contraction before the medial sites of deletion and ellipsis, Selkirk (1984:374ff) makes the plausible proposal that retention of the unreduced auxiliary pre-focus is one of a suite of stylistic metrical options that speakers may use to highlight prosodic and structural parallelism in constructions like those in (21)a–b above.16

In sum, tensed ASYLLABIC contractions are simultaneously prosodified both to the left, as part of the phonological word \((X + I)_\omega\), and to the right, like the tensed weak syllabic auxiliaries, which are metrically dependent on their complement phrase. In contrast, clitics and weak function words that are only leftward-dependent can occur phrase-finally. Compare (24)a with (24)b–d:

(24) a. *Who’s very polite? *Tom’s. (=Tom is)*

b. *Who’s child is very polite? Tom’s. (= Tom’s child)*

c. *Kill ’em. [kIl.ım]*

d. *I might’ve. [’mət’.əv]*

16Inkelas & Zec (1993) postulate an optional syntactic “dislocation” operation that by adjoining the focused phrase to a higher node, places the auxiliary in phrase-final position in the cases where contraction is inhibited.
7 A hybrid model and its consequences

Applied to tensed auxiliary contractions in English, the theory of lexical sharing in LFG simultaneously accounts for their syntactic non-constituency, their lexical morphophonology, and their prosodic and metrical phonology. In effect, by co-instantiating X and I, lexical sharing creates a single lexical and phonological word without necessarily affecting the c-structure or f-structure of X or I. The I retains the rightward metrical dependence of weak syllabic auxiliaries in this position; the phonological word keeps the contraction from being broken up by pauses and strong prosodic boundaries; and the lexical word is the domain of the word-internal phonetic and phonological properties that fuse the host and auxiliary.

But in that theory the hallmarks of usage and lexicalization that contractions bear as lexically shared words are still merely “morphophonological idiosyncracies,” and the theory offers no explanation for why the usage probabilities of adjacent occurrences of words affect both their degree of morphophonological idiosyncrasy and their likelihood of contraction. A straightforward way to improve the theory is to embed it in the mental lexicon of usage-based linguistics (Section 3).

The proposal that lexical sharing belongs to the mental lexicon is visualized in Figure 8. In Figure 8 the “cloud” of memory traces for the contraction you’re is the same as that in Figure 2. The label associated with this cloud is now a structural description, consisting of the lexical entry for the contraction under the theory of lexical sharing. Each cloud of memory traces constitutes a local probability distribution, so the visualization depicts a HYBRID MODEL of usage-based and formal grammar, in which levels of representation are combined with detailed probability distributions learned from experience and constantly updated through life, as proposed by Pierrehumbert (2006) for phonetics.17

As in Section 3, Figure 8 is greatly simplified for intuitive exposition. The key points are that the lexicon stores as allomorphs both the phonologically reduced contracted forms and their usage probabilities, along with any other grammatical, semantic, and pragmatic information that has accrued to them. While Wescoat’s (2005) proposal already lexically stores reduced forms of tensed auxiliary contractions, the current proposal adds representations of their associated usage probabilities/information content. In the exemplar-dynamics model

---

17See also multilevel exemplar-theoretic models of lexical memory (Hay & Bresnan 2006, Walsh et al. 2010).
you’re [juːɪ/jʊɪ/jɔɪ] ← D

(I ↓ pred) = ‘PRO’ (↓ tense) = pres
(↓ pers) = 2
(↓ subj num) = pl
↓ = ↓
(↓ subj) = c ↓

Figure 8: Visualization of tensed auxiliary contractions in a hybrid formal/exemplar model of the mental lexicon

of the mental lexicon, the latter are implicitly represented in the structure of the model, with infrequent and temporally remote memory traces less accessible for lexical insertion into utterance contexts. With the formal LFG grammar extended by lexical sharing, lexical insertion into utterance contexts is formally and computationally explicit. The contracted forms, although non-constituents, are lexical units which are lexically accessed as wholes.

The proposal that the lexical word created by lexical sharing belongs to the mental lexicon makes several direct predictions. First, non-pronouns have more information content than pronouns before the tensed auxiliary is/’s, so on the theory of lexical sharing in the mental lexicon, their likelihood of contractions should be lower. This direct prediction is borne out in variable is contraction data from the Canterbury Corpus of spoken New Zealand English (Gordon et al. 2004), collected by Bresnan & Hay:

---

18 Again, this particular model is one among various alternative implementations of a usage-based lexicon. See Section 3.
19 Because tensed auxiliary contraction with lexical hosts is common only with is/’s and occurrences of has/’s in the dataset are relatively rare, the informativity (average contextual predictability) of the host in this case is proportional to its bigram probability with following is/’s.
20 Jen Hay and the author initially collected and annotated the data in 2015 at the New
Out of 11,719 total observations of variable full and contracted *is*, 88% follow adjacent subject pronouns and 12% follow adjacent non-pronouns.

- Contraction appears with 96% of the former and 43% of the latter observations.

A second prediction is that among non-pronoun hosts before *is’s*, those that have less information content should tend to have higher chances of contraction. This prediction is also borne out by data from the Canterbury Corpus. The non-pronoun hosts having least information content in the Canterbury Corpus *is*-contraction dataset are *one, mum, dad*, and *thing*. These have a far higher proportion of contractions than the average for non-pronouns: = 0.837. Some authentic examples appear in (25):

\[(25) \text{ and my poor } \textit{Mum’s} \text{ here going oh I wish I was there} \]

and *I said come quick come quick*. *Dad’s* at home and he’s a hell of a mess

*one’s* a um . a raving . feminist an *one’s* a chauvinist

*I’ve got [a] friend that has three cats and one’s* a really spiteful cat .

liturgy that they all join in on . and the whole *thing’s* sung

*I wonder if that that kind of* *thing’s* like hereditary

While these data points are suggestive, what is needed to test the consequences of the hybrid theory is a statistical model. After all, there are 1,368 instances of 741 different non-pronoun potential hosts in the dataset, and there are many contributors to *is* contraction other than information content.

8 A statistical model of *is* contraction

To find out whether information content of the hosts affects contraction after controlling for the other predictors, a multiple logistic regression was fit to the data. The model incorporates the following predictors.

Zealand Institute for Language, Brain, and Behavior; research assistant Vicky Watson manually checked them against the audio files for transcription accuracy and marked data exclusions following MacKenzie (2012:65–90).
Information content

The main variable of interest, the information content of the host before is/’s, is calculated as in Section 2. Here, the estimates of bigram and unigram probabilities come from the frequencies of host + is/’s and is/’s in the entire Canterbury Corpus of 1,087,113 words.

Host phrase WC

Host phrase word count (WC) is one of the best predictors of contraction (Frank & Jaeger 2008; MacKenzie 2012, 2013; J. Spencer 2014). WC is a convenient proxy for phrasal weight or complexity, which may make the host phrase more likely to be phrased separately, set off by a phonological or intonational phrase boundary. Table 4 provides authentic examples and Table 5 shows the relation to contraction in the data.

<table>
<thead>
<tr>
<th>host phrases (bolded)</th>
<th>word count</th>
</tr>
</thead>
<tbody>
<tr>
<td>but now work’s just so busy . . .:</td>
<td>WC=1</td>
</tr>
<tr>
<td>the work’s so much harder:</td>
<td>WC=2</td>
</tr>
<tr>
<td>all this blood’s pouring out the side of my head:</td>
<td>WC=3</td>
</tr>
<tr>
<td>some of the work is a bit tedious:</td>
<td>WC=4</td>
</tr>
</tbody>
</table>

Table 4: Host phrase word count (WC)

<table>
<thead>
<tr>
<th>host phrase WC:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4+</th>
</tr>
</thead>
<tbody>
<tr>
<td>proportion contracted:</td>
<td>0.74</td>
<td>0.55</td>
<td>0.43</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Table 5: Proportion contracted by host phrase word count

Year of birth

With non-pronoun hosts, younger speakers of New Zealand English (those born from 1961 to 1987) use contraction more than older (those born from 1926 up to 1961), as Table 6 shows.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>proportion contracted:</td>
<td>0.50</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Table 6: Proportion contracted by speaker year of birth
Class

Nonprofessional NZE speakers use contraction more than professionals (Table 7).

<table>
<thead>
<tr>
<th>class:</th>
<th>N</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>proportion contracted:</td>
<td>0.65</td>
<td>0.49</td>
</tr>
</tbody>
</table>

Table 7: Proportion contracted by speaker class

Previous instance

If the previous instance is *is* or *’s*, the likelihood of *is* contraction is respectively lowered or raised (Table 8). See Szmrecsányi (2005) on “structural persistence.”

<table>
<thead>
<tr>
<th>previous instance:</th>
<th>’s</th>
<th>is</th>
<th>none</th>
</tr>
</thead>
<tbody>
<tr>
<td>proportion contracted:</td>
<td>0.599</td>
<td>0.342</td>
<td>0.510</td>
</tr>
</tbody>
</table>

Table 8: Proportion contracted by previous occurrence of *is/’s*

*is* type

The *is* auxiliary verb contracts more than the *is* copula (cf. Labov 1969; Rickford et al. 1991; McElhinny 1993; MacKenzie 2012; J. Spencer 2014), as Table 9 shows.21

<table>
<thead>
<tr>
<th><em>is</em> type:</th>
<th>aux</th>
<th>cop</th>
</tr>
</thead>
<tbody>
<tr>
<td>proportion contracted:</td>
<td>0.635</td>
<td>0.548</td>
</tr>
</tbody>
</table>

Table 9: Proportion contracted by auxiliary type

Other predictors

Other potential predictors were considered for inclusion: speaker’s gender, whether the final segment of the host is a consonant or vowel, the stress level of

---

21 For a more refined analysis of construction types see Barth (2011) and Barth & Kapatsinski (2017) and also compare MacKenzie’s (2012) discussion of following constituent category.
the final segment, the length of the host in segments, and the number of syllables of the host. All of these added nothing to the model: they had coefficients less than the standard error and were dropped.

In addition, various metrical or prosodic properties of the host phrase were tested as alternatives to WC for another project: (1) total metrical feet (Sternberg et al. 1978, Sternberg et al. 1988); (2) edge boundary strength, manually annotated as the number of lexical word brackets summed with the number of major syntactic phrase (NP, VP, CP) brackets that separate the host from the verb, theoretically corresponding to phonological phrases in Match Theory (Selkirk 2011); (3) cumulative stress from manual annotation of perceived stress values, with and without transformation to a grid format (Liberman & Prince 1977); and (4) cumulative stress based on manually corrected automatic annotation of theoretical stress values, transformed to grid formats. (1) and (4) were automatically annotated using software developed by Anttila et al. (2016). WC substantially improves the model fit compared to alternatives (1) and (4), while (2) and (3) are both competitive with WC. WC is retained here as a convenient proxy until further research is completed.

The fitted model

A multiple logistic regression “working independence” model (Harrell Jr 2001) was constructed from these predictors together with information content of the host before is’s; the numerical predictors were standardized. After the model was fitted to the data, it was corrected for intra-speaker correlations by bootstrap cluster sampling with replacement using the bootcov() function of Harrell Jr (2018). The resulting parameter values are shown in the final fitted model in Table 10.

The model in Table 10 predicts the probability of contraction of any example, given its predictor values. The top line formula converts log odds (used by the regression model) to probabilities. Below, the initial value 0.8804 is the INTERCEPT, representing the overall likelihood of contraction (measured in log odds) when all of the predictor values are zero. The subsequent numerical values are COEFFICIENTS of the model formula, which weight the various predictors and show whether they increase or decrease the overall log odds of a contraction when they do not have zero value; positive coefficients add to the likelihood of the contraction given by the intercept, while negative coefficients reduce the likelihood. The predictors in square brackets are binary valued indicators of categorical properties—professional/nonprofessional class; auxiliary/copula is
\[
\text{Prob}\{\text{Contracted} = 1\} = \frac{1}{1 + e^{-X\hat{\beta}}}, \text{ where}
\]

\[
X\hat{\beta} =
\begin{align*}
& 0.8804 \\
& -0.4741 \times -\log_2 P(\text{host}|\text{verb}) \\
& -0.9868 \times [\text{previous instance} = \text{is}] \\
& -0.2177 \times [\text{previous instance} = \text{none}] \\
& -1.0068 \times \text{host phrase WC} \\
& -0.7060 \times [\text{class} = \text{P}] \\
& -0.5370 \times [\text{is type} = \text{cop}] \\
& +0.4515 \times [\text{year of birth} = [1961, 1987]]
\end{align*}
\]

and \([c] = 1\) if subject is in group \(c\), 0 otherwise

Table 10: Model of Canterbury Corpus variable \(is\) contraction data with non-pronoun hosts

type; speaker year of birth in the earlier or later interval of years. One of the categorical property values is taken to be zero and included in the intercept to calculate the overall likelihood of contraction; when the alternative property value is observed, the overall likelihood is accordingly adjusted by multiplying the coefficient by 1 and adding the result to the total.\(^{22}\) The non-categorical predictors \(-\log_2 P(\text{host}|\text{verb})\) (information content of the host given the verb) and host phrase WC (host phrase word count) have scalar values which are also multiplied by their coefficients. This and similar model formulas are used to validate the model by assessing its predictions on unseen data.

The model quality is reasonably high.\(^{23}\) Partial effects of the model are plotted in Figure 9. The predictors are all reliable within 95% confidence bands,

\(^{22}\)The three-valued predictor for previous instance is decomposed into two binary two-valued predictors: full is vs. 's, and no previous instance vs. 's.

\(^{23}\)Validation of the model found that more than 95% of averaged observed minus expected values in 35 bins are within 2 standard errors (see Gelman & Su’s 2018 binnedplot() function); all predictors have low multicollinearity (condition number \(c < 5\), \(\text{vif} < 1.1\)); average Concordance is \(C > 0.758\) under 10-fold cross-validation with bias correction for speaker clusters in each fold—an “optimism” of < 0.01.
except for the case when the value of previous instance is “none”; there were too few data points for that estimate to be reliable. Because the scalar predictors are standardized, they are plotted on the same scale and the much larger effect of host phrase WC is clearly visible from the greater range it covers on the y-axis. The information content of the host nevertheless has a clear effect as well: greater information content depresses the log odds of contraction.

The author replicated this finding on data from the Buckeye corpus of spoken mid-West American English. The dataset contains 557 instances of variable is-contraction data from 297 different non-pronoun hosts. The predictors are the same except for age and class, which were unavailable or unrelated to contraction in this dataset. Modeling and validation by the same methods as before showed a reliable effect of information load of the host on contraction.

Barth & Kapatsinski (2017:40–41) conducted a multi-model analysis of is/’s contractions with non-pronoun hosts in a smaller dataset of spoken language from the Corpus of Contemporary American English (Davies 2008–). They report that by far the most explanatory predictor among those they used is the bigram probability of host (their “Preceding JP”) and Krug’s (1998) “string frequency”), which is proportional to the information load of the host (n. 19).

9 Restricted vs. unrestricted contractions

The statistical model shows that for variable is contractions with non-pronoun hosts, there is a reliable effect of information content of the host in the context of the verb. The effect is robust after controlling for other contributors to contraction.

It is striking that even the most highly “promiscuous” of the English tensed auxiliary clitics bears the marks of usage probabilities in the effect of information content of the probability of contraction. Restricted and unrestricted contractions have formally similar analyses under the theory of lexical sharing in LFG. They differ in how far they diverge from the uncontracted forms. Restricted contractions diverge more extensively, with morphophonological irregularities and the functional restriction to subject pronouns. But even unrestricted contraction diverges grammatically from the full form where ’s but not is can be used with both singular and plural nouns (Dixon 1977, Nathan 1981, Sparks 1984, Kaisse 1985):

24 The author extracted and annotated the dataset independently of J. Spencer’s (2014) dataset, in order to examine a greater range of host phrase lengths.
Figure 9: Partial effects of the model in Table 10. Each panel shows the effect of one predictor when all of the others are held constant. 95% confidence bands are from the bootstrapped cluster resampling of speakers. The vertical ticks on the plot lines of the numerical covariates (host phrase WC and host information content) show the data densities along the predictor scales.

(26) a. Where’s my pants?
   *Where is my pants?
b. How’s your feet?
   *How is your feet?

c. There’s the lions.
   *There is the lions.

Whether restricted or unrestricted, all tensed auxiliary contractions undergo word-internal morphophonological rules or allomorphy; all tensed aux contractions form a prosodic word with their host to the left and are metrically dependent on the phrase to the right; all asyllabic tensed auxiliaries have some grammatical properties that diverge from their syllabic forms; and for all tensed aux contractions, the likelihood of contraction inversely correlates with the information content of the host in the context of the auxiliary.

Furthermore, the differential selectivity for hosts between restricted and unrestricted contractions is partially eroded. In some varieties of English, some restricted auxiliaries contract with proper noun hosts (Wescoat 2005:471), as in (27) (though in other varieties they do not):

(27)  \textit{Lee’d} seen it. \textipa{li:.@d/li:d}

\textit{Bligh’d} have seen it. \textipa{blaI.@d/blaId}

And restricted contractions do seem to occur with some high-frequency host nouns, such as (Barron 1998:247, n. 13):

(28) \textit{The BBC’ve reported . . .}

— and as in example (29) from the Buckeye corpus:

(29) \ldots their life \textit{people’ve been saying . . .}

10 \textit{I dunno} parallels and implications

It turns out that rather specific properties of restricted auxiliary contractions captured in Wescoat’s (2005) lexical sharing analysis also appear in the multi-word expressions like \textit{I don’t know} studied by Bybee & Scheibman (1999) and Scheibman (2000)—a small but striking parallelism with broader implications.

First, special pronunciations appear only with the most frequent subjects. Bybee & Scheibman (1999:580) observe that in their \textit{don’t} data, though flapping
of [d] occurs only with pronoun subjects, the further reduction of the [o] to [ə], occurs only with the subject I, the most frequent of the pronouns. Likewise, Table 3 illustrates pronunciations of tensed auxiliary contractions specific to the most frequent pronoun subjects, such as I’ll [əl].

Second, Bybee & Scheibman (1999:590) observe that an adverb intervening between the subject and don’t blocks vowel reduction (though it is not blocked by an adverb between don’t and the verb). Likewise, the most reduced pronunciations of the subjects of restricted auxiliary contractions are blocked by an intervening adverb:

\[
\begin{align*}
(30) \text{a. } & \text{I’ll [əl/əl]} \text{ certainly come.} \\
& \text{I [əl/*ə]} \text{ certainly ’ll [əl/*] come.} \\
\text{b. } & \text{They’re [ðei/*də] certainly expensive.} \\
& \text{They [ðei/*də]} \text{ certainly ’re [əl/*] expensive.}
\end{align*}
\]

Third, don’t reduction fails with a conjoined pronoun I and with a lexical subject (Kaisse 1985, Scheibman 2000), as (31)a,b illustrate. (Following Scheibman (2000), the orthographic representation of reduced I don’t know as I dunno is used here.)

\[
\begin{align*}
(31) \text{a. } & \text{*John and I dunno.} \\
\text{b. } & \text{*Those people dunno.}
\end{align*}
\]

The same syntactic restrictions characterize the restricted contractions, as already seen in examples (2) and (6).

The illustrative lexical entries in (32)–(34) are sufficient to capture all three properties of parallelism between contraction and I dunno reduction:25 (1) the dependence on the specific pronoun I for the pronunciation of don’t as [rə], (2) the required adjacency of I and don’t for this reduced pronunciation, and (3) the syntactic restrictions against a conjoined subject with I, (31)a, and against a lexical noun phrase subject, (31)b.

\[
\text{(32) don’t [dət/də] } \leftarrow \text{ I} \\
\quad (\downarrow \text{ TENSE}) = \text{ PRES} \\
\quad (\downarrow \text{ POLARITY}) = \text{ NEG} \\
\quad \neg(\downarrow \text{ SUBJ PERS}) = 3 \\
\quad \downarrow = \downarrow
g\]

\(^{25}\text{Zwicky & Pullum (1983) provide evidence that n’t is an inflectional affix; see also Huddleston & Pullum (2002).}\)
Note that the verb know in (34) is specified intransitive, under the hypothesis that the special pragmatic functions associated with reduction require an unspecified complement. The orthographic rendering I dunno seems to implicate this special pragmatic function. Compare (35)a,b, where the transitive use in (35)b seems less acceptable:

(35)  a.  I dunno, Fred. I’m not sure I agree with you.

b.  ??I dunno Fred. Who is he?

This intransitivity could be the reason for the reported ungrammaticality of examples (35)a,b, discussed by Scheibman (2000) and Kaisse (1985):

(36)  a.  *Tell me what you think I dunno __ well enough.

b.  *The procedure that I dunno __ involves applying to the grad school.

Now consider some implications of these parallels. First, on the traditional view of English tensed auxiliary contraction, it is a purely phonological process of simple cliticization. Phonologists have noted that the asyllabic auxiliaries are PROSODICALLY DEFICIENT, lacking the syllable nucleus of other English words, and they must therefore adjoin to an adjacent prosodic word by phonological cliticization (e.g. Anderson 2008). But this analysis does not extend to don’t, which contains its own syllable nucleus and is not prosodically deficient. Hence, the traditional view of simple cliticization as purely phonological fails to capture its striking parallelisms with the I dunno cases.
Second, the reduced instances of *I don’t know* and the like are **multi-word expressions**. The analysis encapsulated in (33) and (34) shows that the theory of lexical sharing in principle allows the lexicalization of **any strings of words** (collocations) which co-instantiate adjacent part of speech categories. This extension brings LFG with lexical sharing into greater alignment with usage-based grammars such as the variants of Construction Grammar discussed by Bybee (2006), and **data-oriented parsing grammars** including (LFG-DOP), discussed by Bod (2006). The latter stores all constructions, so that the database of constructions (the corpus) *is* the grammar.

Third, the parallels outlined above suggest that what theoretically “triggers” lexical sharing of both constructions like tensed auxiliary contractions and multiword expressions like *I don’t know* is the same: the high probability or predictability of adjacent syntactic elements, just as Bybee and colleagues have argued. It is interesting that the lexical sharing of the small *I dunno* construction—which could be viewed at first glance as a grammatically isolated case—shows its usage-based character to be so similar to the lexical sharing of tensed auxiliary contractions, which are traditionally viewed as a systematic part of English grammar.

Altogether, the present study suggests that combining qualitative and quantitative modeling techniques from both formal grammar and usage-based linguistics could lead to the discovery of usage-based origins in many more regions of syntax.

---

26 Although most previous work with lexical sharing in LFG has concerned contraction, cliticization, and portmanteau-word phenomena with prepositions and determiners (Wescoat 2007, 2009; Broadwell 2008; Alsina 2010; Lowe 2016), Broadwell (2007) already extends the theory to certain multi-word expressions that form phonological words in Zapotec.
References


——, & VSEVOLOD KAPATSINSKI. 2017. A multimodel inference approach to categorical variant choice: Construction, priming and frequency effects on the choice between full and contracted forms of am, are and is. Corpus Linguistics and Linguistic Theory 13.203–260.


