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

(revised December 10, 2019)

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 Avery Andrews, Arto Anttila, Joan Bybee, Janet Pierrehumbert, and Geoff Pullum for providing critical comments on early drafts of the write up. None of these colleagues are responsible for the use I have made of their comments. Thanks also to Jen Hay for her generosity in sharing the ONZE data from the Canterbury Corpus and the resources of the New Zealand Institute for Language, Brain, and Behavior during my visit there as a Fellow in 2015. And thanks to Arto Antilla for many stimulating discussions and collaborations on projects involving corpus phonology and syntax.
At first sight, formal theories of grammar and usage-based linguistics appear completely opposed in their fundamental assumptions. As Diessel (2007) describes it, formal theories adhere to a rigid division between grammar and language use: grammatical structures are independent of their use, grammar is a closed and stable system, and is not affected by pragmatic and psycholinguistic principles involved in language use. Usage-based theories, in contrast, view grammatical structures as emerging from language use and constantly changing through psychological processing.

Yet there are hybrid models of the mental lexicon that combine formal representational and usage-based features, thereby accounting for properties unexplained by either component of the model alone (e.g. Pierrehumbert 2006). Such hybrid models may associate a set of ‘labels’ (for example levels of representation from formal grammar) with memory traces of language use, providing detailed probability distributions learned from experience and constantly updated through life. This study presents evidence for a hybrid model of English tensed auxiliary contractions based on Wescoat’s (2002, 2005) theory of lexical sharing in LFG.

Tensed auxiliary\(^1\) 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 contractions law’s and hell’s in the sentences my other brother in law’s Arab (authentic example from the Buckeye corpus, Pitt et al. 2005) and who the bloddy hell’s knocking (from the Canterbury Corpus, Gordon et al. 2004) are not compositional components in the semantics of the sentences, even though the clitic auxiliary ‘s provides a coda of the open syllables of law and hell that select 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), or *Who’d would you say accept? vs. Who would you say would accept? 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

\(^1\)The 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.

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 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, Bresnan & Spencer 2012, J. Spencer 2014, Barth & Kapatsinski 2017, Barth 2019).

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; Ito & Mester 2018). “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, 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, Barth 2019). Sociolinguistic research on the topic in the Labovian tradition has
generally adopted quantitative methods for modeling variation, as well as the
representational basis of generative grammar, usually with the primary focus
on relating the grammar of the copula to social factors (Labov 1969; McElhinny

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 contrac-
tions of preposition-determiner combinations (witness German zum, am, im,
ins, etc. and French du, au, des, aux, etc.) discussed by Wescoat (2007), and
contractions of simple clitics like English tensed auxiliary contractions (Wescoat
2005), the subject of the present study. However, as the present study shows,
it naturally extends to the lexicalization of multi-word sequences in larger con-
structions. While the original work solved formal representational problems in
theoretical syntax and (as the present study will show) enables an elegant ac-
count of the syntax and prosody of tensed auxiliary contraction, it ignores the
role of usage probabilities in lexicalization, like most work in formal grammar.
On the other hand, usage-based linguistic studies of tensed auxiliary contraction
have seldom presented fully articulated proposals for their syntactic represen-
tations, leaving a wealth of systematic grammatical properties out of account.
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 ex-
plained theoretically. The following three sections outline the main findings
of formal research on tensed auxiliary contraction, and show how the theory
of lexical sharing in LFG improves on previous proposals in capturing these
findings elegantly and accurately. The final three sections first show how the
usage-based and formal approaches naturally fit together into a hybrid model,
next report empirical studies of corpora that support a crucial consequence
of the hybrid theory, and finally extend the theory itself into a usage-based
empirical domain studied by Bybee & Scheibman (1999).
A note on data sources and methods

In keeping with the goal of synthesis, the present study draws on data sources and methods from both formal and usage-based linguistics. The data consist of grammaticality judgments from the linguistic literature and the author’s own speech, as well as authentic evidence from corpora. If an example is not labelled “authentic,” it is constructed. The primary sources of authentic data are (1) the Buckeye Corpus (Pitt et al. 2005) of spoken mid-American English, and (2) the Canterbury Corpus (Gordon et al. 2004) of spoken New Zealand English. Quantitative datasets of variable tensed auxiliary contractions from both corpora are visualized in plots or statistically modeled. In addition, judgment data of examples are validated or corrected, where possible, using examples from Buckeye and Canterbury as well as MacKenzie’s (2012) careful corpus study of auxiliary contraction in spoken English, and finally the web, for a few rarer constructions.

The Buckeye Corpus consists of one-hour interviews with each of 40 people, amounting to about 300,000 words. The speakers are Caucasian, long-time local residents of Columbus, Ohio. The language is unmonitored casual speech. The data are stratified by age and gender: 20 older (defined as age 40 or more), 20 younger; 20 male, 20 female. The words and phones are aligned with sound waves, orthographically transcribed, and provided with broad phonetic labeling.

The Canterbury Corpus is a subcorpus of the Origins of New Zealand English corpora (ONZE). It consists of recorded and orthographically transcribed interviews conducted by students enrolled in a third-year New Zealand English class. Speakers are born between 1930 and 1984, and interviews are added every year with the aim of filling a sample stratified by age, gender, and social class. At the time of collection of the data used in this study, the entire Canterbury Corpus consisted of 1,087,113 words.

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, indicating their lexical storage as separate units from their components. 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></th>
<th>dōt, dō</th>
<th>rōt, rō</th>
<th>rō</th>
<th>ō</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preceding</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<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></td>
<td>14</td>
</tr>
<tr>
<td>we</td>
<td>2</td>
<td>6</td>
<td></td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>they</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>lexical NP</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Following</strong></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>1</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>have</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>have to</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>want</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>see</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td>4</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.\textsuperscript{2}

<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>

Table 2: Full-vowel and reduced-vowel variants of don’t by lexical versus pragmatic function in data from Bybee & Scheibman (1999:587).

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

\textsuperscript{2}Applying a one-sided Fisher exact test to Table 2 to ascertain whether the odds ratio of vowel reduction co-occurring with the pragmatic function is reliably greater than 1, as predicted, yields $p$-value = 0.02545.

6
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).

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, Bresnan & Spencer 2012, J. Spencer 2014, Barth & Kapatsinski 2017, Barth 2019). 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, informativeness*, and related measures.

The measure adopted in the present study is the negative logarithm of the conditional probability of the host given the auxiliary. The conditional prob-

---

3In 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), Bresnan & Spencer (2012), Barth & Kapatsinski (2017), J. Spencer (2014), and Barth (2019) support effects of usage probability on contraction throughout the frequency range.
ability (1a) is estimated as in (1b). The inverse of the conditional probability is its reciprocal, which grows smaller as the probability grows larger and approaches zero as the probability approaches one; inversely, very low probabilities yield extremely high values in the reciprocal. The logarithm of this inverse, which compresses extreme values, yields (1c), here termed the informativeness.\(^4\)

\[
\begin{align*}
(1) \quad & \text{(a) Conditional probability:} \\
& P(\text{word}_1|\text{word}_2) \\
& \text{(b) Estimated:} \\
& \frac{\text{count}(\text{word}_1\text{word}_2)}{\text{count}(\text{word}_2)} \\
& \text{(c) Informativeness:} \\
& \frac{1}{P(\text{word}_1|\text{word}_2)} = -\log P(\text{word}_1|\text{word}_2)
\end{align*}
\]

Why choose a measure of usage probability of the host given the following word, and not the preceding word? One answer is that probability conditioned on the next word can be viewed as measuring the lexical accessibility of \(\text{word}_1\) in the context of the speaker’s planned next word, \(\text{word}_2\); the ratio measures \(\text{word}_1\)’s share of all tokens that precede \(\text{word}_2\); it thus corresponds to \(\text{word}_1\)’s relative salience or activation in that context. The probability of the auxiliary given the preceding word \(P(\text{word}_2|\text{word}_1)\) would presumably be more helpful to the listener, who does not have access to the speaker’s planned next word.\(^5\) Another answer is that conditional probability derived from the following context is often a better predictor than that derived from the preceding context in speech production processing data (Ernestus 2014).

\(^4\)In information theory this quantity is known as the surprisal (Shannon 1948). The information content of a word \(\text{word}_i\) in information theory (Shannon 1948) is the weighted average of its surprisal \(-\log_2 P(\text{word}_i|\text{context}_j)\) over all contexts \(j\); however, in cases of a single specific linguistic context of interest (such as a simple clitic), averaging makes no difference to the value. The present study uses the term “informativeness” in this case. The “informativeness” of a host in the context of an auxiliary is also proportional to its joint probability with the auxiliary—in other words, their probability of occurring together. This latter is Bybee’s (2001, 2010) measure of usage probability.

\(^5\)Several recent studies provide evidence from both corpora and experiments in favor of an accessibility-based model of speech over a model based on uniform information density which could be interpreted as favoring the hearer (Zhan and Levy 2018, 2019).
Figure 1 plots the relation between informativeness and contraction of present tense *have* and *be* with pronominal hosts from the Buckeye Corpus. Figure 1 clearly shows a strong inverse relation between the log likelihood of contraction and informativeness of the pronoun hosts before the verb forms: the first person singular pronoun *I* has the least informativeness before the first person singular verb form *am*, and that sequence has the highest log likelihood of contraction. As informativeness 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**).
Figure 1: Relation between the log odds of contraction and the informativeness of pronoun hosts in the context of verb forms in the Buckeye corpus. The have and be datasets are respectively plotted with magenta and cyan dots, with a loess smoother showing the trend in the combined data.

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.\footnote{The theoretical types of frequency effects generated by the model depend on the parameter range for memory decay and are broader than discussed here. More recent work has developed the exemplar dynamic model further to incorporate the perceptual biases of the listener (Todd 2019, Todd et al. 2019).}

The mental lexicon also stores multi-word fragments that have persistently high usage probabilities (Bybee & Scheibman 1999). These could easily include fragments such as you are, the uncontracted sequence of function words that is functionally equivalent to you’re in grammatical structure. Instances of both
labels: \( \text{you} [\text{j}u:/\text{j}e] \quad \text{you’re} [\text{j}u:\text{r}/\text{j}u:/\text{j}e] \quad \text{are} [\text{a}r/\text{j}e] \)

memory traces:

\[
\begin{array}{cccc}
\text{[jə]} & \text{[jʊɪ]} & \text{[əɪ]} \\
\text{[jə]} & \text{[jə]} & \text{[jʊɪ]} & \text{[əɪ]} & \text{[əɪ]} \\
\text{[jʊɪ]} & \text{[jʊɪ]} & \text{[jʊɪ]} & \text{[əɪ]} & \text{[əɪ]} & \text{[əɪ]} \\
\text{[jʊɪ]} & \text{[jʊɪ]} & \text{[jʊɪ]} & \text{[əɪ]} & \text{[əɪ]} & \text{[əɪ]} & \text{[əɪ]} \\
\text{[jʊɪ]} & \text{[jʊɪ]} & \text{[jʊɪ]} & \text{[əɪ]} & \text{[əɪ]} & \text{[əɪ]} & \text{[əɪ]} \\
\text{[jʊɪ]} & \text{[jʊɪ]} & \text{[jʊɪ]} & \text{[əɪ]} & \text{[əɪ]} & \text{[əɪ]} & \text{[əɪ]} \\
\text{[jʊɪ]} & \text{[jʊɪ]} & \text{[jʊɪ]} & \text{[əɪ]} & \text{[əɪ]} & \text{[əɪ]} & \text{[əɪ]} \\
\end{array}
\]

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).

would have a common label at the level of grammatical labeling. The labeling of new memory instances is based on their similarity to existing instances.

In this way the mental lexicon implicitly encodes bigram probabilities and informativeness as activation levels of the various words and multi-word fragments that are stored there. The crucial connection between high-probability (low-informativeness) host-auxiliary bigrams and higher incidences of contraction in speech production is then straightforward:

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

- 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-informativeness outputs.\(^8\)

If highly probable and relatively uninformative contractions are lexically stored with phonetic detail, they should accrue their phonetic reductions and

---

\(^8\) 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).
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 idiosyncracies” among tensed auxiliary contractions, shown in Table 3. One of them is that “I [aI] may be pronounced [aI], but only in association with ‘ll (will), yielding [aI]; moreover you may become [jɔ], but only when followed by ’re (are), resulting in you’re [jɔ].” Thus the reduced pronunciations are specific to individual pronoun-auxiliary sequences. He emphasizes that these pronunciations are not fast-speech phenomena: I’ll [aI] and you’re [jɔ] “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>I’ll</th>
<th>[aI]</th>
<th>I’m</th>
<th>[am/*am]</th>
<th>I’ve</th>
<th>[arv/*arv]</th>
</tr>
</thead>
<tbody>
<tr>
<td>you’ll</td>
<td>[ju:l/*jɔl]</td>
<td>you’re</td>
<td>[ju:u/*jʊ]</td>
<td>you’ve</td>
<td>[ju:v/*jəv]</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 or darkening of /l/ in will and the laxing of immediately preceding unstressed vowels, yielding we’ll [wi:əl, wIl] in contrast to the rhyming but stressed proper noun the Cree’ll [ˈkɪəl, *kʊIl]). 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 host + auxiliary.

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:

\footnote{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.}
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 informative) words become reduced (shorter).\textsuperscript{10} Seyfarth (2014) discusses this and possible alternative models of the effects of informativeness, 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 variant models of the fundamental usage-based insight connecting lexicalization with probability and reduction.\textsuperscript{11} The exemplar-dynamic model is adopted here for its intuitive expository value, but alternatives could be adopted as well.

4 The grammatical contexts of contraction

Studies of the grammatical contexts that permit or prohibit contraction—particularly their syntax and prosody—have provided the main findings of research on the topic in formal linguistics. Yet, despite its explanatory depth, usage-based linguistics has not provided a detailed understanding of these contexts.\textsuperscript{12} The following three subsections summarize those findings most relevant to the present study. For these it is useful to distinguish between unstressed

\textsuperscript{10}See also Bybee & McClelland (2005) for discussion of a distributed connectionist alternative model, as well as Todd (2019) and Todd \textit{et al.} (2019) for an updated computational model of exemplar dynamics.

\textsuperscript{11}One alternative Seyfarth proposes assumes that only word-specific probabilities and not reduced forms themselves are stored, but that proposal would not very naturally account for the accrual of lexically specific phonetic, semantic and pragmatic accruals of the kind found by Bybee & Scheibman (1999) (see Tables 1 and 2).

\textsuperscript{12}—although Barth and colleagues analyze contraction by broad construction type such as copula, future, and progressive (Barth 2011, Barth & Kapatsinski 2017, Barth 2019).
syllabic and asyllabic forms of the tensed auxiliaries as in Table 4 adapted from Inkelas & Zec (1993) and Wescoat (2005), who follows Sweet (1890:14–16).  

<table>
<thead>
<tr>
<th></th>
<th>full (“strong”)</th>
<th>unstressed syllabic (“weak”)</th>
<th>asyllabic (“enclitic”)</th>
</tr>
</thead>
<tbody>
<tr>
<td>are</td>
<td>[ə]</td>
<td>[ə/ɪ]</td>
<td>[ɪ]</td>
</tr>
<tr>
<td>am</td>
<td>[æm]</td>
<td>[əm]</td>
<td>[m]</td>
</tr>
<tr>
<td>had</td>
<td>[hæd]</td>
<td>[(h)əd]</td>
<td>[d]</td>
</tr>
<tr>
<td>have</td>
<td>[hæv]</td>
<td>[(h)əv]</td>
<td>[v]</td>
</tr>
<tr>
<td>has</td>
<td>[hæz]</td>
<td>[(h)əz]</td>
<td>[z/s]</td>
</tr>
<tr>
<td>is</td>
<td>[ɪz]</td>
<td>[ɪz/əz]</td>
<td>[z/s]</td>
</tr>
<tr>
<td>will</td>
<td>[wɪl]</td>
<td>[əl/l]</td>
<td>[l]</td>
</tr>
<tr>
<td>would</td>
<td>[wʊd]</td>
<td>[(w)əd]</td>
<td>[d]</td>
</tr>
</tbody>
</table>

Table 4: Strong, weak, and enclitic forms of the tensed auxiliaries

**Metrical dependence on the right context**

The asyllabic forms of contracted tensed auxiliaries share metrical constraints on their right contexts with the unstressed syllabic forms of the same auxiliaries. This relation is what Selkirk (1984:405) describes as “the central generalization” about auxiliary contraction: “only auxiliaries that would be realized as stressless in their surface context may appear in contracted form.” It is also the core generalization of Labov’s (1969) analysis, which phonologically derives the asyllabic forms from the syllabic.

The right context of both syllabic and asyllabic reduced auxiliaries requires that the auxiliary be followed by a stressed word, as (2a,b) illustrate.  

(2) a. They are/*’re __ [dɛi ə/l/*dɛi ə/l/*dɛi]  

    b. They are/*’re here. [dɛi ə/l/*dɛi ə/l/*dɛi]  

The stressed word need not be adjacent to the auxiliary. In line with Labov’s (1969) observations, *is* reduces and contracts before the nonadjacent stressed verb *doing* in (3a), but not before unstressed *it* alone.  

---

13In the present study [ə] represents the stressless mid-central vowel and [i] represents a slightly higher unstressed vowel.  
14Following Wescoat (2005), the dot ‘.’ marks a syllable boundary.  
15MacKenzie (2012:80) cites spoken corpus data showing the same effect with several unstressed pronouns.
(3) a. *That bird, what’s it doing ? [w4ts it /dum]/[w4t.ɔ/z it /dum]

b. *That bird, what’s it ? *[w4ts it]/*[w4t.ɔ/z it]

(cf. . . ., what is it?/what’s it?)

Stressed constituents falling outside of the complement phrase of the auxiliaries do not support contraction. In (4a,b) and (5), for example, Labov (1969) and Inkelas & Zec (1993:234) analyze the temporal adverbs as outside the complement phrase of the reduced or contracted *is*:

(4) a. [Tom’s wild] [at night]

b. *[Tom’s ] [at night]

(5) I don’t know where the party is [iz/*iz/*iz] tonight.

Apart from the phrase-final blockage of contraction, the generative literature has repeatedly reported that tensed auxiliary reduction and contraction are blocked in instances of phrase-medial sites of gaps or ellipsis (e.g. Bresnan 1973, Kaisse 1985, Inkelas & Zec 1993, Wilder 1997). For example, Inkelas & Zec (1993) report that the unstressed syllabic forms (as well as the asyllabic contractions) are ungrammatical before instances of “gapping”, “comparative subdeletion”, and “pseudo-gapping”, respectively illustrated by their examples shown in (6a–c):

(6) a. Tom is planting millet, and Lisa is/*iz] peanuts.

b. Karen is a better detective than Ken is/*iz] an archeologist.

c. John is hammering nails with a track hammer but Mary is/*iz] with her shoe.

They argue that this phenomenon is purely a matter of phrasing of the parallel foci with no need for empty syntactic categories.\footnote{For examples of analyses based on empty categories see Anderson (2008) and Ito & Mester (2018). Inkelas and Zec, in contrast, 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.} As evidence they exhibit examples like (7a,b), where the presence of a stressed complement word licenses tensed auxiliary reduction and contraction, despite the presence of an extraction site or empty category following the auxiliary:
(7) a. I don’t know how much there’s/there is [iz/z] __ left in the tank.

b. There’s/there is [iz/z] __ standing on my foot a 300 pound gorilla.

To these we can add attested similar examples from the web showing contractions adjacent to the extraction sites:

(8) 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.

c. No clue what there’s __ going on.

d. ...they were probably aware of what there’s __ going on with her with the fandom.

Selkirk (1984) also suggests that medial sites allow variable contraction in usage, citing (9). 17

(9) Looks as good as it’s __ fun to play (Selkirk 1984:443, n.25)

There are attested examples on the web in support of her contention:

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

b. ...the spherical earth ... shows Australia as being 4 times as as long as it’s wide, ...

c. I still think he’s a better drummer than he’s a singer but don’t tell him that.

d. 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!

17—although MacKenzie (2012:73–74) finds no contractions in the six tokens of comparative subdeletion in her spoken corpus.
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 (9a–b) above.

The main finding important to the present study is that the unstressed tensed auxiliary forms (both syllabic and asyllabic) are metrically dependent on their complement constituents to the right.

Note that there are enclitics and weak function words that are not rightward metrically dependent and hence can occur phrase-finally. Compare the tensed auxiliary in (11a) with (11b–d), a possessive enclitic, a weak object enclitic, and an untensed auxiliary enclitic.

(11) a. _Who’s very polite? *Tom’s. (= Tom is)_  
    b. _Whose child is very polite? Tom’s. (= Tom’s child)_  
    c. _Kill ’em. [kIl.m]_  
    d. _I might’ve. [mAIt.@v]_

**Enclisis with the left context**

While sharing their metrical dependence on a stressed complement in the right context, the unstressed syllabic and asyllabic tensed auxiliaries diverge with respect to the left context.

According to MacKenzie (2012:76–79) contraction of the auxiliary is not found in “pseudo-clefts”, “th-clefts”, and “all-clefts”, respectively illustrated by the authentic examples in (12a–c) from her corpus study:

(12) a. _What I’m talking about is [iz] the people over here and over here and across the street._  
    b. _Well, the problem is [iz], that most of the record players now will not play them._  
    c. _All I know is [iz] I didn’t vote for him._

(12a–c) can be thought of as “specification constructions” in that the post-auxiliary constituent specifies the meaning of the subject, as though preceded
by a colon. The specification appears to form a focused phrase, which can be set off by a pause.

Inkelas & Zec (1993:243,245) propose a phonological explanation that could apply to such specification constructions as well as other authentic pre-auxiliary constructions from MacKenzie (2012)\(^\text{18}\) and constructed examples from Kaisse (1979, 1983, 1985), such as these:

\(\text{(13)}\)
\begin{enumerate}
  \item \textit{Speaking tonight is\textquoteright s our star reporter}.
  \item \textit{That John finally ate is\textquoteright s making Momma happy}.
\end{enumerate}

Inkelas & Zec (1993) assume that English auxiliary clitics form a phonological word \(\omega\) with a phonological word to their left: \(\[ [ \_{\omega} \ _{CL} ]_{\omega} \). 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. How this proposal would apply to (12) and (13) is illustrated in (14).

\(\text{(14)}\)
\begin{enumerate}
  \item *\{ Speaking \ \textbf{tonight} \ \} \{ \text{'s} \_{\omega} \ \text{our star reporter} \}
  \item *\{ That John finally \ \textbf{ate} \ \} \{ \text{'s} \_{\omega} \ \text{making Momma happy} \}
  \item *\{ What I\textquoteright m talking \ \textbf{about} \ \} \{ \text{'s} \_{\omega} \ \text{the people over here . . .} \}
\end{enumerate}

If contractions are enclitics on their hosts to the left, they cannot be interrupted by pauses or by the obligatory prosodic boundaries of certain focused or dislocated syntactic phrases. In contrast, the tensed weak SYLLABIC auxiliaries are not enclitics but 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:

\(\text{(15)}\)
\begin{enumerate}
  \item \{ Speaking \ \textit{tonight} \ \} \{ \text{\textit{is} [iz]} \ \text{our star reporter} \}
  \item \{ That John finally \ \textit{ate} \ \} \{ \text{\textit{is} [iz]} \ \text{making Momma happy} \}
  \item \{ What I\textquoteright m talking \ \textit{about} \ \} \{ \text{\textit{is} [iz]} \ \text{the people over here and over here and across the street} \}
\end{enumerate}

\(^{18}\)These would include parentheticals, adverbs, and preposed prepositional phrases (locative inversions).
They—bicycle cranks, I mean—are [ər] expensive

In sum, tensed ASYLLABIC contractions are simultaneously prosodified both to the left, as part of a phonological word with the host, and to the right, like the tensed weak syllabic auxiliaries, in being metrically dependent on their complement phrases. In other words, both the syllabic and asyllabic forms are ‘proclitic’ in a purely metrical sense (cf. Bresnan 1971, Wilder 1997), but unlike its syllabic counterpart, the asyllabic form also encliticizes to its preceding host.

Restrictive and nonrestrictive auxiliaries

While all asyllabic tensed auxiliary auxiliaries share the properties of metrical dependence on their rightward complements and enclisis on their leftward hosts, further grammatical differences divide them into subtypes that Wescoat (2005) terms RESTRICTIVE and NONRESTRICTIVE. The following classification of asyllabic forms of the tensed auxiliaries is adapted from Wescoat (2005).

(17) restrictive nonrestrictive

<table>
<thead>
<tr>
<th></th>
<th>'re</th>
<th>'d</th>
</tr>
</thead>
<tbody>
<tr>
<td>are</td>
<td>'re</td>
<td>'d</td>
</tr>
<tr>
<td>am</td>
<td>'m</td>
<td></td>
</tr>
<tr>
<td>had</td>
<td></td>
<td>'d</td>
</tr>
<tr>
<td>have</td>
<td>'ve</td>
<td></td>
</tr>
<tr>
<td>has</td>
<td></td>
<td>'s</td>
</tr>
<tr>
<td>is</td>
<td></td>
<td>'s</td>
</tr>
<tr>
<td>will</td>
<td>'ll</td>
<td></td>
</tr>
<tr>
<td>would</td>
<td></td>
<td>'d</td>
</tr>
</tbody>
</table>

The restrictive auxiliaries select pronoun hosts and *wh-* pro-forms, as Wescoat (2005) shows with these examples:

(18) a. I’ll help. [ail]

b. Ai’ll help. [ai*l/*ail]

---

19Wescoat follows Spencer’s (1991: 383) classification of 'd as restrictive, but notes that it contracts with non-pronoun hosts in Zwicky’s (1970) and his own speech, indicating a possible dialectal difference with Spencer’s British English variety. The nonrestrictive classification of 'd is adopted here, because it accords with the author’s variety of American English.

20According to Wescoat (2005), Ai is a Japanese given name and the Au refers to speakers of a language of Papua New Guinea (see Simons & Fennig 2018).
(19) a. We’re a big group. [wiːl]
   
   b. The Cree’re a big group. [kiːr/*kiːl]

(20) a. They’ve gone. [ðeɪv]
   
   b. They may’ve gone. [mei.əv/*meɪv]

(21) a. I’m happy. [ɛm]
   
   b. So am I. [soʊm/*soʊm]

(22) a. How’ve you been? [hauv]
   
   b. The Au’ve been polled. [au.əv/*auv]

In sum, the restrictive auxiliaries generally fail to contract with hosts other than pronouns and wh-pro-forms. Instead, the less reduced weak form that provides its own syllabic peak appears.

The nonrestrictive auxiliaries is, has, had and would can all contract with both pronoun and non-pronoun hosts in some varieties of American English, as the following examples slightly adapted from Wescoat (2005), illustrate.

(23) a. Pat’s gone/going. [paːts]
   
   b. So’s John. [soʊz]

(24) a. She’d seen it. [ʃiːd]
   
   b. Lee’d seen it. [liːd]

(25) a. I’d have seen it. [aɪd]
   
   b. Bligh’d have seen it. [blaɪd]

21 Restrictive contractions do seem to occur with some high-frequency host nouns, such as this example from Barron (1998:247, n. 13)—

The BBC’ve reported . . . [biːbɪ’sɪrv]

—and the following example from the Buckeye corpus:

. . . all their life people’ve been saying. . . [piːplv]
But there are also subtle differences in selectivity for hosts among the non-restrictive auxiliaries. For example, 's for is, has contracts with all wh-proforms, but (at least in the author’s speech) 'd for had, would doesn’t. 'd contracts with who (Who’d like to come with me? and someone who’d failed) but not how, unless it represents inverted did:

(26)  

\begin{align*}
\text{How’s it going? [haʊəz, hauz]} & \quad 's < \text{is} \\
\text{How’s it gone so far? [haʊəz, hauz]} & \quad 's < \text{has} \\
\text{How’d it happen? [hau̯d]} & \quad 'd < \text{did} \\
*\text{How’d it happened? [⁎hau̯d]} & \quad 'd < \text{had} \\
*\text{How’d it have happened? [⁎hau̯d]} & \quad 'd < \text{would}
\end{align*}

Even nonrestrictive 's has a number-neutral use where it selects for a small set of pro-form hosts allowing both singular and plural complement nouns, unlike the full form is (Dixon 1977, Nathan 1981, Sparks 1984, Kaisse 1985):

(27)  

a. Where’s my pants?
   *In what location’s my pants?
   (cf. *Where is my pants?)

b. How’s your feet?
   *In what condition’s your feet?
   (cf. *How is your feet?)

c. There’s the cattle.
   *The cattle’s there.
   (cf. *There is the cattle.)

These data indicate that while the restrictive auxiliaries contrast with the non-restrictive auxiliaries in overwhelmingly restricting their hosts to pronouns and pro-forms, even nonrestrictive auxiliaries show some selectivity for their host.

There is a further syntactic difference between restrictive and nonrestrictive auxiliaries, illustrated by examples (28a–c) from Wescoat (2005): the hosts of the former cannot be conjuncts or occur embedded within a larger subject phrase.

(28)  

a. [She and I] ’ll help. [ar.əl/*ail/*al]

b. [The people beside you] ’re going. [juˌə.1/*juː.1/*jʊ.1/*jɔ.1]

c. [The people who helped you] ’re kind. [juˌə.1/*juː.1/*jʊ.1/*jɔ.1]
In contrast, the following authentic spoken examples (29a,b) from the Canterbury Corpus (Gordon et al. 2004) and (29c,d) from the Buckeye Corpus (Pitt et al. 2005) illustrate that nonrestrictive ’s can contract with noun hosts that are dependents of the subject of the auxiliary and conjuncts:

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

   b. [anything to do with money]’s [z] good

   c. [everybody in my family]’s [z] mechanically inclined.

   d. [August September and October]’s [z] just gorgeous

Other nonrestrictive auxiliaries may not contract as freely as ’s. A single instance of ’d contracted with a non-pronoun host occurs in 2890 occurrences of contracted and uncontracted did, had, would in the Buckeye Corpus:

(30) Wexner Center’d [ˈsɛntʃəd] <SIL> be one of my primary ones.

Judgments of constructed data are uncertain, but (31a–d) suggest that both would and had can contract with a host embedded within a larger subject phrase, at least in the author’s speech:

(31) a. [Everybody in my family]’d agree. [ˈfæməlɪd] d < would

   b. [Everybody in my family]’d agreed to it. [ˈfæməlɪd] d < had

In sum, beyond their shared prosodic and metrical properties, the contracting auxiliaries vary in their selectivity for the host words and differ in their restrictiveness toward host phrases. A number of the enclitic forms show mixed properties, so that the picture that emerges of the restrictive/nonrestrictive classification is a gradient one of overlapping clusters of properties. The most restrictive require that the host be a subject pronoun or wh-pro-form not embedded within a larger subject phrase. The least restrictive lack both of these requirements and very freely encliticize to their adjacent hosts, even nonsubjects.
5 Lexical sharing

Tensed auxiliary contractions, with their morphophonological evidence for host + auxiliary allomorphy, lexical selection of the host, and varying restrictions on host phrases, are problematic for the traditional view of contraction as prosodic enclisis, as Wescoat (2005) argues. When viewed as purely phonological phrasings of two adjacent non-constituent words in the surface syntax, they are not fully accounted for by theories of metrical and prosodic phonology (e.g. Selkirk 1984, 1996; Inkelas & Zec 1993; Anderson 2008; Anttila 2017, Ito & Mester 2018). But Wescoat also argues against lexicalist counteranalyses which propose that the pronoun + restrictive auxiliary contractions have been morphologized into affixed words, for example Sadler’s (1998) LFG analysis and Bender & Sag’s (2001) HPSG analysis, drawing on Spencer (1991:383). The essential problem is that the contractions appear to be morpholexical units but do not also behave like syntactic and semantic units. They cannot be conjoined together, and they permit coordination of the auxiliaries together with their verb phrases:

\[(32) \ a. \quad * [\text{They’re} \text{ and you’re}] \text{ going.} \]
\[ \ b. \quad [\text{I’m looking forward to seeing you}] \text{ and } [\text{will be there on Sunday}] \]
\[ \ c. \quad [\text{You’ll do what I say}] \text{ or } [\text{will suffer the consequences}] \]

The theory of lexical sharing in LFG (Wescoat 2002, 2005) provides a formal analysis of tensed auxiliary contractions in English that solves these problems, 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.\(^{22}\)

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

---

\(^{22}\)The particular category names are not important; here Wescoat follows the c-structure theory outlined by Bresnan (2001) (also Bresnan et al. 2015), but any appropriate category labels will do. The intuition behind D and I is that these are function word categories corresponding to bleached nominals and verbs (Bresnan 2001). In early work, Postal (1966) observes that pronouns behave like determiners in English phrases like we men, you guys, and German anaphoric uses of die, der also support the D analysis of pronouns more generally.
Figure 3: Relation between c-structure and tensed auxiliary contraction under lexical sharing (Wescoat 2005)

Figure 3: Relation between c-structure and tensed auxiliary contraction under lexical sharing (Wescoat 2005)

Wescoat (2005) initially applies the lexical sharing analysis to the restrictive contractions: “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 restrictive contractions are associated with adjacent syntactic terminal categories and may specify itemspecific phonology and functional restrictions, as illustrated in (33). In (33)

---

23 The ‘down’ arrows in (33) 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 (33), 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*
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.

(33) Lexical entries for the structure in Figure 3:

\[\text{you’re \ [ju:ə/ jʊə/jɔɪ]} \leftarrow \begin{array}{ll}
\text{D} & \text{I} \\
(\downarrow \text{PRED}) = \text{‘PRO’} & (\downarrow \text{TENSE}) = \text{PRES} \\
(\downarrow \text{PERS}) = 2 & (\downarrow \text{SUBJ NUM}) = \text{PL} \\
\downarrow = \downarrow & (\downarrow \text{SUBJ}) = c \downarrow
\end{array}\]

\[\text{going \ [ɡʊʊɪŋ]} \leftarrow \begin{array}{l}
\text{V} \\
(\downarrow \text{PRED}) = \text{‘GO\{\downarrow \text{SUBJ}\}\’} \\
(\downarrow \text{ASP}) = \text{PROG} \\
\downarrow = \downarrow
\end{array}\]

Figures 4 and 5 provide extensional visualizations of the structures and relations specified by these lexical entries. The visualization in Figure 4 illustrates that the host must be the subject of the enclitic verb in the functional structure. Figure 5 shows the relations and structures specified by the lexical entry for the verb going in (33). These fragmentary lexical structures are merged and integrated in specific syntactic contexts, such as that in Figure 3.

Wescoat shows that the correct f-structure for Figure 3 follows from general principles of structure-function mapping (Bresnan 2001, 103; Bresnan et al. 2015). These are visualized in Figure 6; 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.)

---

with that of the entire contraction, while the equation \((\downarrow \text{SUBJ}) = c \downarrow\) 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. Note that Wescoat (2005) and Wescoat (2009: 612) adopt different but functionally equivalent formulations; the present analysis follows the latter.

What is missing from this visualization is the constraint (designated by the \(= c \downarrow\) notation in (33)) that in the shared entry the atomic D component f-structure must be the value of the SUBJ function in the atomic I component f-structure. In other words, in the surrounding syntax, whatever it may be, the host must turn out to be the subject of the enclitic verb. This constraint is applied after entire functional description of the utterance is constructed from all of the other equations by the solution algorithm.

---

24
The main prosodic, syntactic, and morphophonological properties shared by all tensed auxiliary contractions follow from this analysis. (1) Host + auxiliary contractions cannot be conjoined to each other as in (32a) because they are not c-structure units. (2) The coordination of two auxiliaries together with their verb phrases, despite the first being contracted with the subject as in (32b,c), is simply I’ coordination, as Wescoat points out. (3) The rightward prosodic dependency of the asyllabic auxiliaries matches those of the weak syllabic forms because they are both stressless auxiliary forms occupying syntactically identical positions on the left edge of their complement phrases. (4) The phonological word status of the host + auxiliary follows from the lexical sharing analysis of tensed auxiliary contractions, given the widely shared assumption of prosodic phonologists that ALL LEXICAL WORDS ARE PHONOLOGICAL WORDS.

As for the properties that distinguish restrictive from nonrestrictive auxiliaries (Section 4), those are captured in the features of the lexical entries making use of the metavariable ↓ (n. 23). In (33), for example, the f-structure of the contraction (which is identified with that of the atomic host pronoun D) must
be the value of the \textit{subj} function of the atomic auxiliary I f-structure. This immediately accounts for syntactic restrictions illustrated in (28a–c), where the host cannot be identified with the subject of the verb because it is only part of the subject.

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 \textit{et al.} 2015):

\[(34) \quad \text{how’ve [hauv]} \leftarrow \begin{array}{cc}
\text{ADV} & \text{C} \\
\downarrow \text{PRED} = \text{‘HOW’} & \downarrow \text{TENSE} = \text{PRES} \\
\downarrow = \downarrow & \downarrow \text{ASPECT} = \text{PERF} \\
& \downarrow \text{FOCUS} = c \downarrow
\end{array}\]

This extension allows restrictive contractions with interrogative pronouns in a parallel way. The lexical entries allow feature selection of the host by the auxiliary.

In this way the theory of lexical sharing improves on preceding theories of restrictive auxiliary contraction.

6 \hspace{1cm} \textbf{Lexical sharing of nonrestrictive contractions}

Wescoat (2005:482) proposes extending the theory of lexical sharing from re-
stricture contractions of tensed auxiliaries to the nonrestrictive tensed auxiliary 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/az] (is or has) to a host, yielding a lexical-sharing structure; the host may be anything, the attachment of ’s [z/s/az] 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. (35) shows the schematic form of lexical entries of contractions of ’s. It differs from the entry for you’re shown in (33) in that here the restriction (↓ subj) = c ↓ is absent and the host and its category are unspecified. This schema can be viewed as the “lexical process” for attaching nonrestrictive ’s to hosts.

(35) Schematic form of lexical entries of contractions of ’s:

\[
x’s [ \ldots z/s/iz ] \leftarrow X I \\
\ldots (\downarrow \text{TENSE}) = \text{PRES} \\
\downarrow = \downarrow (\downarrow \text{SUBJ NUM}) = \text{SG} \\
(\downarrow \text{SUBJ PERS}) = 3
\]

The choice of the specific pronunciation of ’s in (35) 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:

(36) 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

\footnote{The generalization to contractions of inverted ’s would allow C as an extended head as well as I; see discussion of (34).}
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:

\[
\begin{align*}
(37) \quad & \text{Pete sang (} [s] \text{) and Fred sang (} [s/*z] \text{)} \\
& \quad \text{Fred zigged (} [z] \text{) and Pete zagged (} [z/*s] \text{)} \\
& \quad \text{Ross zig-zagged (} [z/*@z] \text{)}
\end{align*}
\]

An example of 's contraction under lexical sharing is given in Figure 7, and the lexical entry of the contraction blood’s is given in (38). Note that the lexical entry has the schematic structure in (35), which requires adjacency in c-structure between the host and auxiliary categories.

Figure 8 shows how the structure in Figure 7 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.

(38) Lexical entry for the contraction blood’s in Figure 7, derived from the schema (35).

\[
\begin{align*}
\text{blood’s [bladz]} & \leftarrow \text{N} \\
\quad (\downarrow \text{PRED}) = \text{‘BLOOD’} \\
\downarrow = \downarrow \\
\text{I} & \leftarrow \text{N} \\
\quad (\downarrow \text{TENSE}) = \text{PRES} \\
\quad (\downarrow \text{SUBJ NUM}) = \text{SG} \\
\quad (\downarrow \text{SUBJ PERS}) = 3
\end{align*}
\]
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 (39)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:

(39) a. I’ll tell you what I don’t think’s going on. [TINks]
b. What I don’t think’s beautiful is a boy in my daughter’s bedroom. [TINks]
c. You can’t oppose what you don’t know’s happening. [nooz]

As (40) and Figure 9 show, the lexical sharing analysis of these cases is straightforward.

(40) Lexical entry for the contraction think’s:

\[
\text{think’s } [\text{TINks}] \leftarrow \begin{array}{c}
V \\
(\downarrow \text{PRED}) = \text{‘THINK((SUBJ) (COMP))’} \\
(\downarrow \text{TENSE}) = \text{PRES} \\
(\downarrow \text{SUBJ NUM}) = \text{SG} \\
(\downarrow \text{SUBJ PERS}) = 3
\end{array}
\]
Notice that the lexical entry in (40) is the same in schematic form as that in (38), even though the resulting grammatical relations between the host noun and auxiliary are entirely reversed. To see the reversal, compare Figure 8, where the host heads a subject which is an argument of the main clause co-headed by the tensed auxiliary, to Figure 9, 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 nonrestrictive contractions in (35) and follow from the general principles of structure-function mapping invoked by Wescoat (2005).

Furthermore, since is contractions are not c-structure constituents under lexical sharing, there is no danger of unwanted ‘movements’ in the lexical shar-

**Between restrictive and nonrestrictive**

As shown in Section 4, the restrictive/nonrestrictive classification of tensed auxiliaries is not a binary classification, but a gradient one of overlapping clusters of properties. How does the lexical sharing analysis account for the intermediate auxiliaries between restrictive and nonrestrictive?

As observed in Section 4, nonrestrictive ’s has a number-neutral use illustrated in (27a–c). These have shared lexical entries similar to (41):

\[
\begin{align*}
(41) \quad \text{how’s [həʊz]} & \leftarrow \text{ADV} \quad \text{C} \\
& \downarrow \quad \text{(↓ PRED) = ‘HOW’} \quad \text{(↓ TENSE) = PRES} \\
& \downarrow = \downarrow \quad \text{(↓ SUBJ PERS) = 3} \\
& \quad \text{(↓ FOCUS) = c} \downarrow
\end{align*}
\]

Unlike the syllabic auxiliary forms, the asyllabic auxiliary specifies person but not the number of the subject and it selects for specific pro-forms as hosts how, yielding How’s your feet? vs. *In what condition’s your feet? and *How is your feet?

As also noted in Section 4, the asyllabic auxiliary ’d is restrictive in some varieties and nonrestrictive in others. A restrictive entry for the conditional mood sense of ’d requiring a pronoun subject is shown in (42):^26

\[
\begin{align*}
(42) \quad \text{I’d [aɪd]} & \leftarrow \text{D} \quad \text{I} \\
& \downarrow \quad \text{(↓ PRED) = ‘PRO’} \quad \text{(↓ MOOD) = COND} \\
& \downarrow = \downarrow \quad \text{(↓ SUBJ PERS) = 1} \\
& \quad \text{(↓ SUBJ) = c} \downarrow
\end{align*}
\]

In this variety Bligh’d have seen it pronounced [*blaId] is ungrammatical. In other varieties the pronoun specification on the host is dropped and the subject [blaId] is fine:

\[
\begin{align*}
(43) \quad \text{x [..d]} & \leftarrow \text{X} \quad \text{I} \\
& \downarrow = \downarrow \quad \text{(↓ MOOD) = COND} \\
& \quad \text{(↓ SUBJ) = c} \downarrow
\end{align*}
\]

---

^26Recall from (26) that past perfect and conditional uses of ’d differ in host selectivity from the past tense use.
Because the host f-structure must be identified with that of the subject of the auxiliary, this shared entry rules out the contraction of conditional or past-perfect ‘d with adverbs, as in examples like So’d Ann *[soud] for So would Ann, as well as accounting for *How’d it have happened? in (26).

For the present author ‘d is even less restrictive, allowing contractions not only with a subject as in (43), but with a dependent of the subject: witness family’d in (31b). The greater degree of contraction is permitted by the lexical entry in (44):

\[
(44) \quad x \ldots d \leftarrow X \downarrow = \downarrow \quad (\downarrow \text{MOOD}) = \text{COND} \quad (\downarrow \text{SUBJ GF*}) = c \downarrow
\]

The same shared entry rules out *Who would you say ’d accept? because the host say of contracted ‘d is not a dependent of the subject of would.

These successful analyses of the gradient contraction phenomena between restrictive and nonrestrictive further motivate the lexical sharing theory for all tensed auxiliary contractions.

7 A hybrid model

The last three sections have shown that the main prosodic, syntactic, and morphophonological properties common to all tensed auxiliary contractions have an elegant analysis under lexical sharing in LFG and that the theory can also accurately characterize the overlapping clusters of fine-grained lexico-syntactic properties of auxiliaries falling between restrictive and nonrestrictive. These formal analyses reveal both the intricacy and systematicity of English grammar in this domain, qualities that the representational tools of formal grammar are well-suited to express. Usage-based accounts have not offered fully articulated representations of the grammatical contexts of contractions that can account for these rich syntactic and prosodic properties, and without detailed formal syntactic representations it is difficult to see how they can do so.\(^{28}\)

\(^{27}\)The notation GF* specifies a possibly empty chain of nested grammatical functions, allowing nonlocal dependencies between the auxiliary’s subject and its host. For this and other details of the formalism see Börjars et al. (2019) or Dalrymple et al. (2019).

\(^{28}\)One variety of formal usage-based grammar is DATA-ORIENTED PARSING grammar including LFG-DOP, discussed by Bod (2006). The latter stores the c- and f-structures of all constructions, so that the database of constructions (the corpus) is the grammar. The the-
The fundamental contribution of the theory of lexical sharing to this account is to admit non-constituent syntactic fragments as lexical units whose parts are still integrated with the larger contextual syntactic and metrical structure. They show the typical effects of lexicalization: lexically stored as wholes, the host + auxiliary sequences may diverge in their morphophonological and other properties from their uncontracted counterparts. As shown in Section 5, previous accounts do not capture all these phenomena. Yet even in the lexical sharing theory these hallmarks of usage and lexicalization are nevertheless merely “morphophonological idiosyncracies.” It is the theory of lexicalization in usage-based linguistics that explains why the usage probabilities of adjacent occurrences of words affect both their degree of morphophonological idiosyncrasy and their likelihood of contraction.

A way to improve both theories is to combine them, and a straightforward way of doing so is to embed the formal theory of lexical sharing in LFG within the mental lexicon of usage-based linguistics (Section 3). LFG is particularly compatible with usage-based linguistics because, as noted above, it is a lexical syntactic theory which maps the “surface” syntactic groupings of words—represented by the c-structures—into a parallel functional structure that models the “deeper” relations and dependencies. Specifications of the global functional structure are locally distributed across the individual words of expressions via their lexical entries, making it easy to represent the FRAGMENTABILITY OF LANGUAGE, as Bresnan (2001) terms one of the consequences of the design principles of LFG: “the fact that we can infer the internal structural relations of arbitrary sentence fragments as easily as we can those of complete sentences, and the supporting fact that sentence fragments are so extensively used” (Bresnan 2001, Bresnan et al. 2015, 79).

Functional structures can correspond to equivalence classes of surface syntactic realizations, both within a language and typologically. For example, you’re going and you are going have equivalent f-structures on the present account: compare the grammatical structure for the former (Figure 6) to that for the latter, shown in Figure 10. Omitting the VP from both structures yields, by the same solution algorithm used for complete sentences, identical f-structures of the fragments, lacking only the pred attribute of the main verb go.29 What

29Several recent books provide introductions to the LFG solution algorithm, which derives f-structures from c-structures and their lexical entries: see Bresnan et al. (2015), Börjars et al.
this equivalence means is that the lexical choice between contracted \textit{you’re} and uncontracted \textit{you are} depends not on syntax, but on phonology, prosody and style. It also means that memory traces of functionally equivalent multi-word fragments like \textit{you’re} and \textit{you are} can be labeled by their common f-structure in the hybrid model.

![Diagram](image)

Figure 10: C-structure to f-structure mapping for \textit{you are going}. Compare to (6).

Multi-word fragments that consist of function words only (auxiliaries, pronouns, and pro-forms) are a finite set of highly frequent elements that speakers can enumerate from memory. It is plausible that these lexically shared function words and their functionally equivalent uncontracted fragments both occur in the mental lexicon. Both fit into c-structures in the same way.

Is it as plausible to hypothesize that nonrestrictive contractions also belong in the mental lexicon together with their equivalent uncontracted fragments? For subject pronouns undergoing is contraction (such as \textit{he, she, it, there, that}) the contractions and their equivalent uncontracted expressions can be represented in the mental lexicon in the same way as the restrictive contractions. For non-pronoun hosts, those with the greatest usage probabilities in the context of \textit{is/’s} or \textit{would/’d, had/’d} plausibly have persistent memory traces, but for many other contractions memory traces would presumably rapidly decay. For these evanescent contractions the lexical schemata (35) for tensed ’s and

\cite{WeSc05, DaSc19}. Wescoat (2005) details how lexical sharing integrates with the LFG solution algorithm.
those like (43), (44) for conditional ‘d representing would could serve to generate short-term shared entries and their functionally equivalent fragments. In general, both the formal grammar labels and schemas that appear in the shared lexical entries represent dimensions along which similarity-based generalization could take place, allowing new instances of utterances never encountered before to be perceived and produced as is contractions, together with their full prosodic and syntactic properties.

8 Empirical studies

Empirical studies have found contractions with pronoun hosts to be a function of usage probabilities—as their storage in the mental lexicon predicts (Section 3). What about contractions with content-word hosts such as blood’s, think’s or in some varieties Bligh’d, Center’d (recall examples (38), (40), (25b), (30))? In accordance with the hybrid model, the mental lexicon may also include memory traces of ‘s and ‘d contractions with content-word hosts, both persistent and evanescent, and they too should be a function of usage probabilities. The present section provides evidence in support of this consequence of the hybrid model, in line with previous studies discussed below.

If the mental lexicon includes nonrestrictive contractions with non-pronoun hosts, then there should be an inverse relation between the proportions contracted of non-pronoun hosts and their informativeness, similar to that shown in Figure 1 for pronoun hosts. To investigate this theoretical prediction of the hybrid model of tensed auxiliary contraction, is contraction data were collected from two spoken corpora, the Buckeye Corpus (Pitt et al. 2005) and the Canterbury Corpus (Gordon et al. 2004). (Nonrestrictive ‘d contractions are set aside because they are much sparser; recall the discussion of (30).) The Canterbury Corpus is over three times as large as the Buckeye Corpus and is annotated for the social variable of class that could play a role in contraction (cf. Labov 1969, McElhinny 1993).30 Hence, it became the main focus of the investigation, with the Buckeye Corpus used to replicate the findings from the Canterbury Corpus.

All instances of is and orthographic ‘s were collected from the Canterbury Corpus transcriptions in 2015 at the New Zealand Institute for Language, Brain, and Behavior; research assistant Vicky Watson manually checked a sample against the audio files for transcription accuracy and also marked data ex-

30The Canterbury Corpus has been used primarily for sociophonetic studies and previous studies of auxiliary contraction in NZ English are lacking.
clusions. Exclusions included hosts with final sibilants (which do not occur with the asyllabic auxiliary 's), instances of 's representing has, and the variety of other grammatical contexts found by MacKenzie (2012:65–90) to be outside the envelope of variation. The hosts in this dataset were labeled as pronouns or non-pronouns, and informativeness was calculated as in (1) from ngram statistics provided by Jen Hay and Robert Fromont for the entire Canterbury Corpus. This yielded 11,719 total observations from 412 speakers (mean instances per speaker = 28, standard deviation = 23) and 758 unique non-pronoun hosts.

For the Buckeye Corpus replications a dataset of variable is contractions was extracted and annotated following a similar method to that of Bresnan & Spencer (2012) and J. Spencer (2014). The author labeled the hosts of this dataset as pronouns or non-pronouns, and calculated informativeness as in (1) from ngram statistics compiled from the entire Buckeye Corpus. After exclusions, there were 4019 total observations from all 40 speakers (mean instances per speaker = 100, standard deviation = 46) and 306 unique non-pronoun hosts.

First, non-pronouns have higher informativeness than pronouns before the tensed auxiliary is/'s, so on the theory of lexical sharing in the mental lexicon, their likelihood of is contractions should be lower. The data bear out this expectation:

- 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.

Comparable data from the smaller Buckeye Corpus of spoken mid-American English show a similar pattern:

- Out of 4019 total observations of variable full and contracted is, 85% follow adjacent subject pronouns and 15% follow adjacent non-pronouns.

---

31 Bresnan & Spencer (2012) and J. Spencer (2014) already show an effect on contraction of (respectively) the log and negative log conditional probability of non-pronoun hosts given is/'s in data collected from the Buckeye Corpus. The present dataset was constructed independently of the datasets described in those studies in order to examine a greater range of host phrase lengths.

32 This observation immediately follows from the fact that pronoun-is/'s bigrams are far more frequent than non-pronoun-is/'s bigrams and the definition of informativeness in (1), given n. 4.
Contraction appears with 92% of the former and 37% of the latter observations.

Secondly, among non-pronoun hosts before is/’s, those that have lower informativeness should tend to have higher chances of contraction. This expectation is also borne out by data from the Canterbury Corpus. The non-pronoun hosts having lowest informativeness 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 (45):

(45) and my poor Mum’s 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

In the Buckeye Corpus is-contraction dataset, the proportion contracted of the least informative non-pronoun hosts (everybody, one, think, everything, daughter, and mom) differs less extremely from the other hosts, but shows a tendency in the predicted direction: 0.453 vs. 0.354.

These simple descriptive statistics support this crucial consequence of the hybrid theory: that is contraction with non-pronoun hosts should show evidence of the probabilistic structure of the mental lexicon. But while these data points are suggestive, what is needed to test this crucial consequence is a statistical model that controls for other possible predictors of contraction. After all, there are many hosts in the dataset, and the literature on contraction has identified many contributors to is contraction other than informativeness (see below). To this end, the author planned a multiple logistic regression model and annotated the corpus data for the variables described below, using the statistical computing platform R (R Core Team 2019) as well as direct inspection and manual annotation of extensive data samples.
Informativeness

The main variable of interest, the informativeness of the non-pronoun 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; Bresnan & Spencer 2012; J. Spencer 2014). WC can be viewed as 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. It could also be viewed as a proxy for phrasal informativeness, in that longer phrases are likely to be more informative in a qualitative sense. Table 5 provides authentic examples and Table 6 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><em>but now work</em>’s just so busy . . .:*</td>
<td>WC=1</td>
</tr>
<tr>
<td><em>the work</em>’s so much harder:</td>
<td>WC=2</td>
</tr>
<tr>
<td><em>all this blood</em>’s pouring out the side of my head:*</td>
<td>WC=3</td>
</tr>
<tr>
<td><em>some of the work</em> is a bit tedious:*</td>
<td>WC=4</td>
</tr>
</tbody>
</table>

Table 5: 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>total instances:</td>
<td>543</td>
<td>500</td>
<td>182</td>
<td>143</td>
</tr>
<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 6: Proportion contracted by host phrase word count

---


34However, quantitative measures of phrasal informativeness run up against the problem of sparseness of data. Even restricting host phrase length to two words, for example, one finds that 90% of the 500 two-word phrases occur just once in the dataset.
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 7 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 7: Proportion contracted by speaker year of birth

Speaker year of birth is numerical data available in the corpus, but it is severely bimodal around the year 1961, causing model fit problems. The year of birth data is therefore dichotomized at 1961.

Class

Nonprofessional NZE speakers use contraction more than professionals (Table 8).

<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 8: 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 9). 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 9: Proportion contracted by previous occurrence of is/’s

Successive instances of is/’s are from the same speaker, and are collapsed across the copula/auxiliary types (see below).
is type

Those instances of is/’s in construction with a participial form of the verb are defined as ‘auxiliaries’, while those in construction with nominals, prepositions, and adjectives are defined as ‘copulas’. 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 10 shows.35

<table>
<thead>
<tr>
<th>is 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 10: 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 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. Interactions were not included because of the complexity of the model in relation to the data.

In addition, various metrical or prosodic properties of the hostphrase 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. (To appear). 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.

35 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 fitted model

Because speaker identity is a source of unknown dependencies in the data, a multiple logistic regression “working independence” model (Harrell Jr 2001) was constructed from these variables, with the numerical variables 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 11.

\[
\text{Prob}\{\text{Contracted} = 1\} = \frac{1}{1 + e^{-X\hat{\beta}}}, \quad \text{where}
\]

\[
X\hat{\beta} =
\]

\[
0.8804
\]

\[
-0.4741 \times -\log_2 P(\text{host}|\text{verb})
\]

\[
-0.9868 \times [\text{previous instance} = \text{is}]
\]

\[
-0.2177 \times [\text{previous instance} = \text{none}]
\]

\[
-1.0068 \times \text{host phrase WC}
\]

\[
-0.7060 \times [\text{class} = \text{P}]
\]

\[
-0.5370 \times [\text{is type} = \text{cop}]
\]

\[
+0.4515 \times [\text{year of birth} = [1961, 1987]]
\]

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

Table 11: Model of Canterbury Corpus variable is contraction data with non-pronoun hosts

---

\(^{36}\)Here the working independence model starts from the assumption that speakers’ utterances are independent of speaker identity, and then corrects this assumption by estimating the extent of these dependencies using bootstrap resampling with replacement of entire clusters (each speaker defines a ‘cluster’ of utterances). Bresnan et al. (2007) describe cluster resampling in this way: “In other words, we can create multiple copies of the data by resampling from the speakers. The same speakers’ data can randomly occur many times in each copy. We repeatedly re-fit the model to these copies of the data and used the average regression coefficients of the re-fits to correct the original estimates for intra-speaker correlations. If the differences among speakers are large, they will outweigh the common responses and the findings of [the working independence model] will no longer be significant.”
The model in Table 11 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 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.\textsuperscript{37} The non-categorical predictors $-\log_2 P(\text{host}|\text{verb})$ (informativeness 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.\textsuperscript{38} Partial effects of the model are plotted in Figure 11. The predictors are all reliable within 95% confidence bands, 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 informativeness of the host nevertheless has a clear effect as well: greater informativeness depresses the log odds of contraction.

The author replicated this finding on the non-pronoun host data from the Buckeye corpus of spoken mid-West American English. The predictors are the same except for age and class, which were unavailable or unrelated to

\textsuperscript{37}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.

\textsuperscript{38}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 $c < 5$, 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 $< 0.01$. 

43
Figure 11: Partial effects of the model in Table 11. 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 informativeness) show the data densities along the predictor scales.

contraction in this dataset. Modeling and validation by the same methods as before showed a reliable effect of informativeness 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 informativeness of the host (n. 4).

In sum, this crucial consequence of the hybrid theory has been borne out by empirical studies of two spoken English corpora in the present study and is buttressed by a further empirical study of a third corpus (Barth & Kapatsinski 2017): is contraction with non-pronoun hosts shows clear evidence of the probabilistic structure of the mental lexicon.

9 I dunno parallels and implications

It turns out that rather specific properties of restrictive auxiliary contractions captured in Wescoat’s (2005) lexical sharing analysis also appear in the multi-word expressions like 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 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 restrictive auxiliary contractions are blocked by an intervening adverb:


b. They’re [ðeI/ðeI] certainly expensive.
    They [ðeI/ðeI] certainly ’re [əl/ə] expensive.

Third, don’t reduction fails with a conjoined pronoun I and with a lexical subject (Kaisse 1985, Scheibman 2000), as (47)a,b illustrate. (Following
Scheibman (2000), the orthographic representation of reduced *I don’t know* as *I dunno* is used here.

(47) a. *John and I dunno.*
    b. *Those people dunno.*

The same syntactic restrictions characterize the restrictive contractions, as already seen in examples (28a–c).

The illustrative lexical entries in (48)–(50) are sufficient to capture all three properties of parallelism between contraction and *I dunno* reduction:39 (1) the dependence on the specific pronoun *I* for the pronunciation of *don’t* as [r5], (2) the required adjacency of *I* and *don’t* for this reduced pronunciation, and (3) the syntactic restrictions against a conjoined subject with *I*, (47)a, and against a lexical noun phrase subject, (47)b.

(48) *don’t* [dɒt/də] ←

\[
\text{D} \quad \text{I}
\]

\[
(\downarrow \text{TENSE}) = \text{PRES} \\
(\downarrow \text{POLARITY}) = \text{NEG} \\
\neg(\downarrow \text{SUBJ PERS}) = 3 \\
\downarrow = \downarrow
\]

(49) *I don’t* [airô] ←

\[
\text{D} \quad \text{I} \\

(\downarrow \text{PRED}) = \text{‘PRO’} \\
(\downarrow \text{PERS}) = 1 \\
(\downarrow \text{NUM}) = \text{SG} \\
\downarrow = \downarrow
\]

\[
(\downarrow \text{TENSE}) = \text{PRES} \\
(\downarrow \text{POLARITY}) = \text{NEG} \\
\neg(\downarrow \text{SUBJ PERS}) = 3 \\
(\downarrow \text{SUBJ}) = \text{c} \downarrow
\]

(50) *I don’t know* [airônoʊ] ←

\[
\text{D} \quad \text{I} \quad \text{V}
\]

\[
(\downarrow \text{PRED}) = \text{‘PRO’} \\
(\downarrow \text{PERS}) = 1 \\
(\downarrow \text{NUM}) = \text{SG} \\
\downarrow = \downarrow
\]

\[
(\downarrow \text{TENSE}) = \text{PRES} \\
(\downarrow \text{POLARITY}) = \text{NEG} \\
\neg(\downarrow \text{SUBJ PERS}) = 3 \\
(\downarrow \text{SUBJ}) = \text{c} \downarrow
\]

\[
(\downarrow \text{PRED}) = \text{‘KNOW(\downarrow \text{SUBJ})’} \\
(\downarrow \text{SUBJ}) = \text{c} \downarrow
\]

Note that the verb *know* in (50) is specified intransitive, under the hypothesis that the special pragmatic functions associated with reduction require an

\[\text{Zwicky & Pullum (1983) provide evidence that } n’t \text{ is an inflectional affix; see also Huddleston & Pullum (2002).}\]
unspecified complement. The orthographic rendering *I dunno* seems to implic- cate this special pragmatic function. Compare (51)a,b, where the transitive use in (51)b seems less acceptable:

(51) 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 (51)a,b, discussed by Scheibman (2000) and Kaisse (1985):

(52) 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 (49) and (50) 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 analysis extends LFG with lexical sharing from the quasi-morphological domain of portmanteau words and simple clitics squarely into the multi-word territory of usage-based linguistics.40

Third, the parallels outlined above suggest that what theoretically “triggers” lexical sharing in both constructions like tensed auxiliary contractions and multiword expressions like *I don’t know* is the same: the high usage probability 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

40 Broadwell (2007) already extends the theory of lexical sharing in LFG to certain multi-word expressions that form phonological words in Zapotec.
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.

From these considerations it is clear that the hybrid model improves on both of its separate components. The formal grammar component provides parallel analyses of English tensed auxiliary contraction and the *I dunno* construction from Bybee & Scheibman’s (1999) classic usage-based study; the formal generalizations revealed extend the coverage of the usage-based theory to what are thought of as more systematic parts of English grammar. The usage-based component provides a theory of lexicalization explaining the “idiosyncractic” properties of restrictive contractions and it correctly predicts an inverse relation between the nonrestrictive contractions with content-word hosts and their probability of adjacent occurrence in usage. Hybridization strengthens both components.

Certainly other varieties of formal grammar could be adopted in hybrid models of the type proposed here. The justification for using LFG with lexical sharing as the formal representational basis is simply that LFG is a probabilistically oriented\(^{41}\) lexical syntactic theory of multi-word fragments, and the incorporation of lexical sharing adds multi-word fragments to its lexicon, supporting its embedding into the mental lexicon of usage-based linguistics. Another theory with similar attributes could serve as well. The contribution of the present work is to provide evidence for extending hybrid models of formal and usage-based linguistics from the empirical domain of the word (as in Pierrehumbert 2006, for example) to the domain of syntax.

References


ANTTILA, ARTO. 2017. Stress, phrasing, and auxiliary contraction in English. In *The Morpho-Syntax Phonology Connection*, ed. by Vera Gribanova &

\(^{41}\)LFG has been adopted as representational basis in many varieties of probabilistically oriented theories of language, from data-oriented processing (Bod 2006) to stochastic optimality theory (Bresnan *et al.* 2001, Wescoat 2007, Maslova 2007) to random fields (Johnson & Riezler 2002). Some early models of statistical NLP were dual-structure models like LFG with dependency graphs labeled by grammatical functions replacing f-structures (de Marneff & Manning 2008).


——. 2019. Effects of average and specific context probability on reduction of function words BE and HAVE. Linguistics Vanguard 5. Published online by DeGruyter, 2019-05-08 DOI: https://doi.org/10.1515.

——, & 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.


